

Molecular genetic approaches to decrease misincorporation of non-canonical branched chain amino acids into a recombinant protein in *Escherichia coli*

Ángel Córcoles García - Dissertation

**Molecular genetic approaches to decrease mis-incorporation of
non-canonical branched chain amino acids into a recombinant
protein in *Escherichia coli***

vorgelegt von

M. Sc.

Ángel Córcoles García

ORCID: 0000-0001-9300-5780

von der Fakultät III-Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Juri Rappsilber, Institut für Biotechnologie, TU Berlin, Berlin

Gutachter: Prof. Dr. Peter Neubauer, Institut für Biotechnologie, TU Berlin, Berlin

Gutachter: Prof. Dr. Pau Ferrer, Universitat Autònoma de Barcelona, Bellaterra
(Cerdanyola del Vallès), Spain

Gutachter: Dr. Heinrich Decker, Sanofi-Aventis Deutschland GmbH, Frankfurt am Main

Tag der wissenschaftlichen Aussprache: 11. Dezember 2019

Berlin 2020

Abstract

The incorporation of non-canonical branched chain amino acids (ncBCAA) such as norleucine, norvaline and β -methylnorleucine into recombinant proteins during *E.coli* production processes has become a crucial matter of contention in the pharmaceutical industry, since such mis-incorporation can lead to production of altered proteins, having non optimal characteristics. Hence, a need exists for novel strategies valuable for preventing the mis-incorporation of ncBCAA into recombinant proteins.

This work presents the development of novel *E. coli* K-12 BW25113 strain mutants allowing exogenous tunable expression of target genes involved in the BCAA biosynthetic pathway. For that purpose, following single genes were knocked out by homologous recombination in order to eliminate endogenous genetic expression: *thrA*, *ilvA*, *leuA*, *ilvIH*, *ilvBN*, *ilvGM* and *ilvC*. Expression regulation of previously knocked-out genes was carried out by transforming arabinose-based tunable expression plasmids containing the native sequence of target genes into the generated *E. coli* KO mutants. The engineered *E. coli* hosts were screened in a mini-reactor system in fed-batch mode under both standard cultivation conditions and under conditions reproducing large-scale effects. During screening three different concentrations of L-arabinose were tested for each strain in order to trigger different expression levels of a target gene. Screening was performed by comparing the impurity profile of the recombinant mini-proinsulin expressed in each tested strain with the non-engineered *E. coli* host. After screening, potential *E. coli* mutants showing the most significant ncBCAA reduction were scaled-up to a 15L stirred tank bioreactor and analysed to confirm their behaviour.

Screening experiments carried out in the mini-reactor system indicated that an up-regulation of *ilvC*, *ilvIH* and *ilvGM* and down-regulation of *leuA* and *ilvBN* trigger a reduction of norvaline and norleucine biosynthesis and mis-incorporation into mini-proinsulin. Results for *ilvA* and *thrA* suggest that the threonine pathway is not the one redirecting more metabolic flux to α -ketobutyrate. Moreover, it is noteworthy that norleucine was the most mis-incorporated ncBCAA and β -methylnorleucine levels did not significantly change under tested experimental conditions. Furthermore, a novel cultivation strategy consisting of pyruvate pulsing and oxygen limitation was demonstrated to successfully mimic large-scale effects since biosynthesis of ncBCAA and metabolites of the overflow metabolism were reported under those conditions. Among the tested genes, up-regulation of *ilvIH* and *ilvGM* showed the highest reduction of ncBCAA biosynthesis and mis-incorporation. Reproducibility of potential *ilvIH* and *ilvGM*-tunable *E. coli* strains was confirmed in the 15L reactor. These novel *E. coli* strains with a reduced ncBCAA mis-incorporation may have a crucial role in the pharmaceutical industry since product quality is pivotal for commercial approval of recombinant proteins that are to be used as human therapeutics.

Zusammenfassung

Der Fehleinbau von nichtkanonischen verzweigtkettigen Aminosäuren (ncBCAA) wie Norleucin, Norvalin und β -Methylnorleucin in rekombinante Proteine während *E. coli* - Proteinproduktionsprozessen ist in der pharmazeutischen Industrie zu einem entscheidenden Kriterium geworden, da ein solcher Fehleinbau zur Produktion veränderter rekombinanter Proteine mit suboptimalen Eigenschaften führen kann. Daher besteht ein Bedarf an neuen Strategien, um den Fehleinbau von ncBCAA in rekombinante Proteine zu verhindern.

Diese Arbeit beinhaltet die Entwicklung von neuen *E. coli* K-12 BW25113-Mutantstämmen, in denen die Expression von Zielgenen, die am BCAA-Biosyntheseweg beteiligt sind, extern gesteuert werden kann. Zu diesem Zweck wurden folgende Gene mittels homologer Rekombination inaktiviert: *thrA*, *ilvA*, *leuA*, *ilvIH*, *ilvBN*, *ilvGM* und *ilvC*. Gleichzeitig wurden diese Mutanten durch Plasmide komplementiert, die das chromosomal inaktivierte Gen unter Kontrolle eines regulierbaren Arabinose-Promoters enthalten. Diese *E. coli*-Stämme wurden in einem Mini-Bioreaktorsystem im Fed-batch-Modus sowohl unter Standardkultivierungsbedingungen als auch im Scale-down-Simulator mit pulsbasierter Zufütterung untersucht. Während des Screenings wurden für jeden Stamm drei verschiedene Konzentrationen von L-Arabinose getestet, um unterschiedliche Expressionsniveaus eines Zielgens auszulösen. Für jeden Stamm wurde dann der Fehleinbau nichtkanonischer Aminosäuren in rekombinant exprimiertes Mini-Proinsulins quantifiziert und mit dem nicht-manipulierten *E. coli*-Kontrollstamm verglichen. Nach dem Screening wurden *E. coli*-Mutanten mit einem signifikant verminderten Fehleinbau zur Validierung der Ergebnisse im Laborreaktor unter Scale-down Bedingungen getestet.

Bei dem im Mini-Reaktorsystem durchgeführten Screening zeigte eine Erhöhung der Expression von *ilvC*, *ilvIH* bzw. *ilvGM* sowie eine Verminderung der Expression von *leuA* bzw. *ilvBN* eine Reduktion der Norvalin- und Norleucin-Biosynthese und einen geringeren Fehleinbau in das Mini-Proinsulin. Die Änderung der Expression der Gene *ilvA* und *thrA* zeigten keinen signifikanten Einfluss. Dies bestätigt die Hypothese, dass die Synthese der nichtkanonischen Aminosäuren nicht über den Threoninweg, sondern durch die direkte Kettenverlängerung von Pyruvat verursacht wird.

Außerdem, ist es bemerkenswert, dass Norleucin die am häufigsten falsch eingebaute nichtkanonische Aminosäure war und sich die β -Methylnorleucin-Spiegel unter getesteten Versuchsbedingungen nicht signifikant änderten. Darüber hinaus konnte gezeigt werden, dass eine neuartige Kultivierungsstrategie, die auf Pyruvat-Pulsen bei gleichzeitiger Sauerstofflimitierung beruht, Scale-down-Effekte erfolgreich imitieren kann. Von den getesteten Zielgenen zeigte eine Expressionssteigerung von *ilvIH* bzw. *ilvGM* den signifikantesten Effekt bezüglich einer verminderten Synthese der nichtkanonischen Aminosäuren und dem Fehleinbau. Diese neuartigen *E. coli*-Stämme, die einen verminderten Fehleinbau nichtkanonischer Aminosäuren zur Folge haben, können eine entscheidende Rolle in der pharmazeutischen Industrie spielen, da die Produktqualität für die Zulassung rekombinanter Proteine als Humantherapeutika von entscheidender Bedeutung ist.

List of Contents

Abstract	I
Zusammenfassung.....	II
List of Contents.....	III
Acknowledgements	VIII
List of Abbreviations and Symbols	X
Abbreviations.....	X
Symbols.....	XIII
1. Introduction	1
2. Literature Review	3
2.1 <i>E. coli</i> as industrial biofactory	3
2.1.1 Advantages of <i>E. coli</i> as industrial biofactory.....	3
2.1.2 Origin and genetic background of strain <i>E. coli</i> K-12 BW25113.....	3
2.1.3 Inclusion body-based recombinant protein expression in <i>E. coli</i>	4
2.1.4 Insulin expression in <i>E. coli</i>	5
2.1.5 Physiological stress during recombinant protein expression in <i>E. coli</i>	7
2.2 Introduction to central <i>E. coli</i> metabolism	8
2.2.1 Catabolite repression and oxidative respiration	8
2.2.2 Mixed-acid fermentation	9
2.2.3 Overflow metabolism.....	10
2.2.4 BCAA biosynthetic pathway	12
2.2.5 Regulation of the BCAA biosynthetic pathway	15
2.3 Non-canonical branched chain amino acids (ncBCAA)	23
2.3.1 Introduction to ncBCAA.....	23
2.3.2 ncBCAA biosynthetic pathway	25
2.3.3 Substrate promiscuity of enzymes involved in ncBCAA biosynthesis	27
2.3.4 Conditions triggering ncBCAA biosynthesis	29
2.3.5 Mechanism of ncBCAA mis-incorporation into proteins during translation.....	31
2.3.6 Inconvenients of ncBCAA biosynthesis and mis-incorporation into recombinant proteins.....	33
2.3.7 Strategies to avoid ncBCAA mis-incorporation into recombinant proteins.....	34
2.4 Recombineering in <i>E. coli</i>	38
2.4.1 Homologous recombination based on FRT/FLP system.....	39
2.5 Tunable promoter systems in <i>E. coli</i> metabolic engineering.....	41
2.5.1 Introductory considerations.....	41

2.5.2	<i>araC</i> -P _{BAD} promoter system.....	42
2.5.3	<i>rhaBAD</i> promoter system.....	43
2.5.4	<i>lac</i> promoter system	44
2.5.5	Pm/XylS promoter system.....	46
2.5.6	Summary	46
3.	Research Hypotheses and Aim of the Project.....	47
4.	Materials & Methods	51
4.1	Materials, reagents and equipment.....	51
4.2	Software	51
4.3	Bacterial strains.....	51
4.4	Plasmids	51
4.5	Media	51
4.6	Primers	52
4.7	Standard molecular biology methods.....	52
4.7.1	PCR.....	52
4.7.2	Plasmid extraction (mini and midiprep).....	54
4.7.3	Genomic DNA extraction.....	55
4.7.4	Preparation of <i>E. coli</i> electro-competent cells.....	55
4.7.5	Electroporation of DNA into <i>E. coli</i> electro-competent cells	56
4.7.6	Curation of temperature-sensitive plasmids.....	56
4.7.7	Electrophoresis analytical gel.....	57
4.7.8	PCR spin purification	57
4.7.9	Extraction from DNA gels	57
4.7.10	Standard cloning based on restriction digestion and ligation.....	57
4.7.11	InFusion cloning.....	58
4.7.12	Mutagenesis	59
4.7.13	Sequencing	59
4.7.14	Preparation of bacterial glycerol stocks.....	60
4.7.15	SDS-PAGE.....	60
4.8	Amino acid analysis by GC-FID	61
4.8.1	Intracellular soluble protein fraction and inclusion body isolation from cell extracts	61
4.8.2	Hydrolysis of intracellular soluble protein fraction and inclusion bodies.....	62
4.8.3	ncBCAA analysis with EZ:faast TM for free (physiological) amino acid analysis by GC-FID .	62
4.8.4	ncBCAA analysis by gas chromatography-flame ionization detection (GC-FID).....	63
4.8.5	Data analysis of data generated by GC-FID	63

4.9	Generation of pSW3 plasmid variants with different LacI expression levels: pSW3, pSW3_ <i>lacI</i> ⁺ and pSW3_ <i>lacI</i> ^q	64
4.9.1	Sequencing of pSW3.....	64
4.9.2	Cloning <i>lacI</i> ^q into pSW3 to generate plasmid pSW3_ <i>lacI</i> ^q	64
4.9.3	Mutation of pSW3_ <i>lacI</i> ^q to generate pSW3_ <i>lacI</i> ⁺	65
4.9.4	Transformation of pSW3 into <i>E. coli</i> K-12 BW25113 and molecular verification	66
4.9.5	Transformation of pSW3_ <i>lacI</i> ^q into <i>E. coli</i> K-12 BW25113 and molecular verification	66
4.9.6	Transformation of pSW3_ <i>lacI</i> ⁺ into <i>E. coli</i> K-12 BW25113 and molecular verification	66
4.10	Design of an arabinose-based tunable expression plasmid (pACG_araBAD)	67
4.10.1	Features of the DNA fragments involved in InFusion cloning	67
4.10.2	Generation of PCR fragments for InFusion cloning.....	67
4.10.3	InFusion cloning to generate plasmid pACG_araBAD	68
4.10.4	Cloning of genes into plasmid pACG_araBAD	69
4.11	Design of an m-toluene-based tunable expression plasmid (pACG_XylSPm).....	71
4.11.1	Generation of plasmid pACG_XylSPm	71
4.11.2	Cloning of genes into plasmid pACG_XylSPm	72
4.12	Development of strains to regulate expression of the BCAA biosynthetic genes (<i>geneX</i> -tunable <i>E. coli</i>)	73
4.12.1	Generation of strains <i>E. coli</i> K-12 BW25113 Δ <i>geneX</i> (<i>geneX</i> : <i>leuA</i> , <i>thrA</i> , <i>ilvA</i> and <i>ilvC</i>).....	74
4.12.2	Generation of strains <i>E. coli</i> K-12 BW25113 Δ <i>ilvIH</i> and <i>E. coli</i> K-12 BW25113 Δ <i>ilvBN</i>	76
4.12.3	Transformation of pSW3_ <i>lacI</i> ⁺ into <i>E. coli</i> K-12 BW25113 Δ <i>geneX</i>	78
4.12.4	Transformation of pACG_araBAD_ <i>geneX</i> into <i>E. coli</i> K-12 BW25113 Δ <i>geneX</i> pSW3_ <i>lacI</i> ⁺ (<i>geneX</i> -tunable <i>E. coli</i>).....	78
4.12.5	Transformation of pACG_XylSPm_ <i>geneX</i> into <i>E. coli</i> K-12 BW25113 Δ <i>geneX</i> pSW3_ <i>lacI</i> ⁺	79
5.	Results	80
5.1	Analysis of mini-proinsulin expression in <i>E. coli</i> K-12 W3110M and <i>E. coli</i> K-12 BW25113 containing plasmid pSW3.....	80
5.1.1	Cultivation conditions.....	80
5.1.2	Mini-proinsulin analysis with SDS-PAGE	81
5.2	Analysis of mini-proinsulin expression in <i>E. coli</i> K-12 BW25113 containing different variants of plasmid pSW3.....	82
5.2.1	Cultivation conditions.....	82
5.2.2	Mini-proinsulin analysis with SDS-PAGE	82
5.3	Evaluation of L-arabinose induction in <i>E. coli</i> BW25113 Δ <i>geneX</i> expressing pSW3_ <i>lacI</i> ⁺ and pACG_araBAD_ <i>geneX</i> (<i>geneX</i> -tunable <i>E. coli</i>)	84
5.3.1	Cultivation conditions.....	84
5.3.2	Gene expression analysis	84

5.4	Evaluation of m-toluate induction in <i>E. coli</i> BW25113 $\Delta geneX$ expressing pSW3_ <i>lacI</i> ⁺ and pACG_XylSPm_ <i>geneX</i> (<i>geneX</i> -tunable <i>E. coli</i>).....	86
5.4.1	Cultivation conditions.....	86
5.4.2	Gene expression analysis	86
5.5	Establishment of a GC-FID method allowing analysis of canonical and non-canonical amino acids	87
5.5.1	Elaboration of calibration curves and determination of retention times for ncBCAA norvaline, norleucine and β -methylnorleucine.....	88
5.5.2	Elaboration of calibration curves and determination of retention times for canonical amino acids.....	88
5.5.3	Evaluation of the effect of hydrolysis on amino acid analysis	90
5.5.4	Validation of the GC-FID method by analyzing a pure protein	91
5.6	Establishment of cultivation conditions based on pyruvate pulsing leading to an increase of ncBCAA mis-incorporation into recombinant mini-proinsulin expressed in <i>E. coli</i> at shake flask level	92
5.6.1	Cultivation conditions.....	93
5.6.2	Analysis of ncBCAA.....	93
5.7	Establishment of cultivation conditions based on pyruvate pulsing and O ₂ limitation leading to an increase of ncBCAA mis-incorporation into recombinant mini-proinsulin expressed in <i>E. coli</i> in a 10 mL mini-reactor and in a 15L reactor	95
5.7.1	Experiment in a 10 mL mini-reactor.....	95
5.7.2	Experiment in a 15 L reactor	102
5.7.3	Comparison 10 mL mini-reactor and 15 L reactor	111
5.8	Screening of <i>geneX</i> -tunable <i>E. coli</i> strains a mini-reactor system	113
5.8.1	Cultivation conditions.....	113
5.8.2	Mini-proinsulin analysis by SDS-PAGE.....	116
5.8.3	ncBCAA analysis.....	116
5.9	Screening of potential <i>ilvGM</i> - and <i>ilvIH</i> -tunable <i>E. coli</i> strains in a 15 L reactor under conditions triggering ncBCAA formation	121
5.9.1	Cultivation mode	121
5.9.2	Mini-proinsulin analysis by HPLC.....	125
5.9.3	Analysis of ncBCAA.....	126
5.9.4	Acetate and formate analysis.....	128
6.	Discussion.....	130
6.1	Analysis of mini-proinsulin expression in <i>E. coli</i> K-12 BW25113 containing different variants of plasmid pSW3.....	130
6.2	Evaluation of L-arabinose induction in <i>E. coli</i> BW25113 $\Delta geneX$ expressing pSW3_ <i>lacI</i> ⁺ and pACG_araBAD_ <i>geneX</i> (<i>geneX</i> -tunable <i>E. coli</i>)	131

6.3	Evaluation of m-toluate induction in <i>E. coli</i> BW25113 Δ <i>geneX</i> expressing pSW3_ <i>lacI</i> ⁺ and pACG_XylSPm_ <i>geneX</i> (<i>geneX</i> -tunable <i>E. coli</i>).....	132
6.4	Establishment of a GC-FID method allowing analysis of canonical and non-canonical amino acids	133
6.5	Establishment of cultivation conditions based on pyruvate pulsing and O ₂ limitation leading to an increase of ncBCAA mis-incorporation into recombinant mini-proinsulin in <i>E. coli</i>	134
6.6	Screening of <i>geneX</i> -tunable <i>E. coli</i> strains	137
7.	Conclusions and Outlook	141
8.	References.....	143
9.	Theses	160
10.	Appendix	161

Acknowledgements

In the following lines I would like to show my appreciation to all people who contributed to the successful accomplishment of this work.

Firstly, I want to express my sincere gratitude to my supervisors Prof. Dr. Peter Neubauer and Dr. Peter Hauptmann for giving me the opportunity to perform my PhD at Sanofi-Aventis Deutschland GmbH in close collaboration with Technische Universität Berlin (TUB). Their supervision, advice and encouragement marked a milestone in how I perceive science nowadays. In the same vein, I would like to thank Dr. Heinrich Decker and Dr. Sebastian Rissom for counting on me to develop my thesis at Sanofi-Aventis Deutschland GmbH.

Secondly, I would like to specially thank all the colleagues from Sanofi-Aventis Deutschland GmbH, who also contributed in various ways to this thesis. I thank Dr. Berndt Janocha, Klaus Sparwald and Michaela Michel for supporting me concerning the use of GC-FID to perform amino acid analysis. I thank Dr. Holger Penders, Dr. Manuel Krewinkel, Claudia Erhard, Frank Ludwig, Alexander Leske, Jose Gonzalez Rubio, Thomas Trick and Melanie Hänig for their supervision and advice concerning fermentations at reactor scale. I thank Dr. Simon Stammen, Dr. Hans-Falk Rasser, Karima Chaouki, Sylvia Zink, Nicole Schottstedt, Anika Ludwig, Martina Hildebrand and Birgit Antonijs for all their help concerning molecular biology. I would also like to express my gratitude to Janine Neumann and Christiane Straub for helping me with the nightmare of the business trips and all the bureaucracy. I also thank Waltraud König for inviting me to the regular lunch meeting for Spanish speakers, which brought me the opportunity to feel a bit more home in a foreign country. In addition, I thank to Reiner Olliger and Thomas Kuhl the possibility to perform HPLC analysis at their laboratories. I would like to thank Sarah Charaf for her help regarding generation of strains *E. coli* K-12 BW25113 $\Delta ilvBN$ and *E. coli* K-12 BW25113 *ilvG+* as well as plasmid pACG_araBAD_ *ilvBN*. It was a pleasure to supervise the development of her master thesis. But I want to particularly show my gratitude to Dr. Peter Hauptmann and the colleagues Klaus Schaffer, Andre Urban and Dietmar Heep. It was a pleasure to share lab with them for 3 years in such a privileged working environment. Their encouragement, advice and sense of humor were one of the main pillars contributing to my motivation and the success of this work.

Thirdly, I would like to express my gratitude to all the colleagues from Technische Universität Berlin (TUB), who also contributed in various ways to this work. I specially thank to Emmanuel Anane, who was my main guide during my secondment at TUB. It was a pleasure to learn more about *E. coli* metabolism, fermentation technology and scale-down models after our interesting discussions. I am happy that our collaboration could translate in a scientific publication. In the same way, I would like to express my gratitude to Dr. Christian Reitz for showing me the homologous recombination technology used in this thesis to perform gene knock outs. My next thanks goes to Prof. Dr. Peter Neubauer, Dr. Stefan Junne, Benjamin Haby, Sebastian Hans and Florian Gauche for their scientific inputs, discussions and experimental support during my secondment at TUB.

Forthly, I would also like to show my appreciation to the colleagues from Denmark Technical University (DTU) who guided me during my secondment there, especially to Prof. Dr. Krist V. Gernaey and Dr. Christian Bach.

I am grateful for financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 643056 (Biorapid). In this regard, I would like to thank the European Union institutions for making it possible for young researchers to live such a rewarding experience. In the same way, I also express my gratitude to all ESRs taking part in the Biorapid Consortium and to the project coordinator, Prof. Dr. Jarka Glassey, for the pleasant moments we spent together during the training weeks and conferences celebrated throughout the project.

Finally, I would like to thank my family and friends in Spain and all over the world who were always encouraging and supporting me despite the distance. I also thank to all my friends in Frankfurt, my cats and, specially, to Mario Hiebel, for his patience and persistent support during the hard times of the PhD life. Thank you all for always being there and trust on me. You were my main motivation to success in this work and do not give up. I love you! ¡Muchas gracias, os quiero mucho!

List of Abbreviations and Symbols

Abbreviations

etc.	et cetera
i.e.	id est
e.g.	exempli gratia
(d)NTP	(deoxy)nucleoside triphosphate
2D	2-dimensional
3D	3-dimensional
AA	amino acid
aaRS	aminoacyl-tRNA synthetase
ABA	L-2-aminobutyric acid
ACKA	acetate kinase
ACS	acetyl-CoA synthetase
ADH	alcohol dehydrogenase
ADP	adenosine diphosphate
AHAS	acetohydroxy acid synthase
AK	bifunctional aspartokinase/homoserine dehydrogenase 1
Ala	alanine
Amp	ampicillin
ASAD	aspartate-semialdehyde dehydrogenase
Asn	asparagine
Asp	aspartic acid (aspartate)
asRNA	antisense RNA
ATP	adenosine triphosphate
BCAA	branched-chain amino acids
bST	bovine somatotropine
CaM	mammalian calmodulin
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
cBCAA	canonical branched-chain amino acids
CDW	cell dry weight
CGSC	<i>E. coli</i> Genetic Stock Center
CISY	citrate synthase
cm, chlor	chloramphenicol
CNBr	cyanogen bromide
CNCl	cyanogen chloride
CoA (CoA-SH)	coenzyme A
CRP	cAMP receptor protein
DH	dihydroxy-acid dehydratase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DO	dissolved oxygen

dsDNA	double-stranded DNA
EDTA	ethylenediaminetetraacetic acid
EF-Tu	elongation factor Tu
EI	enzyme I
EII	enzyme II
ER	endoplasmid reticulum
FAD	flavin adenine dinucleotide in oxidized form
FADH ₂	flavin adenine dinucleotide in reduced form
FdhF	cytoplasmic molybdenum- and selenium-dependent formate dehydrogenase H
FHL	formate hydrogenlyase complex
FID	flame ionization detection
FLP	site-specific flipase recombinase
FRT	FLP recombinase recognition sites
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GC	gas chromatography
GDP	guanosine diphosphate
Gln	glutamine
GLU	glutamate
Glu	glutamic acid (glutamate)
Gly	glycine
Gpp	pppGpp-5'-phosphohydrolase
GTP	guanosine triphosphate
His	histidine
hM-CSF	human macrophage colony stimulating factor
HPr	histidine-containing protein
HSAT	homoserine acyltransferase
HSD	bifunctional aspartokinase/homoserine dehydrogenase 1
HSK	homoserine kinase
Hyc	hydrogenase component
IB	inclusion body
IHF	integration host factor
IL-2	interleukin-2
Ile	isoleucine
ileRS	isoleucyl-tRNA synthetase
IPMD	3-isopropylmalate dehydrogenase
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	ketol-acid reductoisomerase (NADP(+))
ISOM	3-isopropylmalate dehydratase
Kan	kanamycin
kanR	kanamycin-resistant
kanS	kanamycin-sensitive
KB	ketobutyrate
KGLU	α-ketoglutarate
KIV	α-ketoisovalerate
KO	knock-out
KV	ketovalerate

L-ara	L-arabinose
LB	Luria-Bertani
LDH	lactate dehydrogenase
Leu	leucine
leuRS	leucyl-tRNA synthetase
lim.	limitation
Lrp	leucine-responsive protein
Lys	lysine
MCS	multicloning site
Met	methionine
metRS	methionyl-tRNA synthetase
NAD(P)+	nicotinamide adenine dinucleotide (phosphate) in oxidized form
NAD(P)H	nicotinamide adenine dinucleotide (phosphate) in reduced form
ncBCAA	non-canonical branched-chain amino acids
Ndk	nucleoside diphosphate kinase
NL	norleucine
Nva	norvaline
OD	optical density
OD _{600nm}	optical density at 600 nm wavelength
PCR	polymerase chain reaction
PDHC	pyruvate dehydrogenase complex
PEP	phosphoenolpyruvate
PFL	pyruvate-formate lyase
PFR	plug flow reactor
Phe	phenylalanine
PI	pro-insulin
PMS, IPMS	α -isopropylmalate synthase
ppGpp	guanosine 5'-diphosphate, 3'-diphosphate
PPI	pre-pro-insulin
pppGpp	guanosine 5'-triphosphate, 3'-diphosphate
pro	proline
PTA	phosphate acetyltransferase
PTS	PEP-dependent carbohydrate phosphotransferase system
PYR-O ₂	pyruvate pulsing and oxygen limitation
dH ₂ O	deionized water
RE	restriction enzyme
Rel. q.	relative quantity
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosomal ribonucleic acid
RSH	protein pair RelA/SpoT
SAM	S-adenosylmethionine
SD, STD	standard
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
Ser	serine
SOC	Super Optimal broth with Catabolite repression

ssDNA	single-stranded DNA
β-MNL	β-methylnorleucine
STR	stirred tank reactor
Ta	annealing temperature
TD	threonine deaminase, encoded by <i>ilvA</i>
Thr	threonine
Tm	melting temperature
TMG	methyl-1-thio-β-d-galactopyranoside
TrB	transaminase B
TrC	transaminase C
tRNA	transference ribonucleic acid
Trp	tryptophane
TS	threonine synthase
TUB	Technische Universität Berlin
Tyr	tyrosine
UAS	upstream activating sequence
Val	valine
WT	wild type

Symbols

μ_{set}	set-point of the specific cell growth rate (h^{-1})
F	feed flow rate over time (L h^{-1})
q_s	specific substrate uptake rate ($\text{gS gX}^{-1} \text{h}^{-1}$)
S	concentration of glucose (g/L)
t	time (h)
V	volume (L)
X	biomass concentration (g/L)
σ^{32}	heat-shock sigma factor 32
σ^s	sigma regulator

1. Introduction

Biosynthesis and mis-incorporation of ncBCAA such as norleucine, norvaline and β -methylnorleucine into recombinant proteins has been reported in a number of *E. coli* production processes. ncBCAA can be synthesized by *E. coli* metabolism as byproducts of the BCAA biosynthetic pathway as a result of the low specificity of the *leu* and *ilv*-operon-encoded enzymes for their substrates. There are three main causes triggering ncBCAA biosynthesis in *E. coli*: overflow metabolism driven by inefficient mixing in large-scale reactors, de-regulation of enzymes encoded by the *leu* operon resulting from leucine depletion and deficiency of AHAS II activity. Free ncBCAA can be mis-incorporated into cellular proteins through tRNA misaminoacylation during protein translation due to their structural similarity with the respective canonical amino acids. Such mis-incorporation might lead to the production of altered recombinant proteins, having non optimal characteristics e.g. altered biological activity, modulated sensitivity to proteolysis and immunogenicity. This represents an important concern for the pharmaceutical industry since product quality is pivotal for commercial approval of recombinant proteins that are to be used as human therapeutics in order to ensure patient safety. Several valuable strategies for reducing mis-incorporation of ncBCAA into recombinant proteins were already described in the literature: mutation of methionine codons of the gene encoding the recombinant protein, co-expression of enzymes capable of degrading ncBCAA, supplementation of exogenous canonical amino acids, overproduction of methionine by mutating genes involved in methionine and threonine biosynthesis, knocking-out genes involved in the BCAA biosynthetic pathway, optimization of fermentation conditions, supplementation of trace elements and use of alternative *E. coli* expression strains less prone to non-canonical BCAA mis-incorporation. However, most of the aforementioned strategies present numerous disadvantages. This thesis focuses on another approach which has not yet been explored and that might contribute to close the scientific gap: engineering of novel *E. coli* strains allowing tunable expression of target genes involved in the BCAA biosynthetic pathway.

Together with the novel Pm/XylS promoter, the classic *araBAD* promoter was also utilized in this thesis in order to regulate expression of target genes involved in the BCAA biosynthetic pathway. In order to ensure functionality of the *araBAD* promoter, strain *E. coli* K-12 BW25113 was selected as the model organism for this study, since it is deficient in arabinose catabolizing enzymes. Furthermore, in order to properly evaluate the effect that expression regulation of a target gene has on ncBCAA biosynthesis and subsequent mis-incorporation into cellular proteins, mini-proinsulin was selected as the model recombinant protein in this study. Protein sequence of recombinant mini-proinsulin contains canonical amino acids (3 methionine, 14 leucine and 5 isoleucine residues) that can be potentially substituted by the non-canonical counterparts (norleucine, norvaline and β -methylnorleucine, respectively) upon mistranslation. An available plasmid (pSW3) expressing recombinant mini-proinsulin under the control of a P_{tac} promoter was tested in the genetic background of *E. coli* K-12 BW25113, revealing leaky and weak expression. Thus, a number of pSW3 plasmid variants encoding for different levels of LacI repressor were tested in *E. coli* K-12 BW25113, being plasmid pSW3_*lacI*^r the variant reporting the most optimal results. Furthermore, a method allowing amino acid analysis (including ncBCAA) by gas chromatography (GC-FID) was successfully established. In order to regulate expression of target genes involved in the BCAA biosynthetic pathway it was first necessary to eliminate endogenous expression of such genes in the *E. coli* cell.

Hence, following single target genes/operons were knocked out from the *E. coli* K-12 BW25113 genome by homologous recombination: *thrA*, *ilvA*, *leuA*, *ilvIH*, *ilvBN*, *ilvGM* and *ilvC*. Expression regulation of previously knocked-out genes was carried out by transforming arabinose and m-toluate-based tunable expression plasmids (pACG_araBAD and pACG_XylSPm series) containing the native sequence of the target genes into the respective engineered *E. coli* K-12 BW25113 KO mutants. Since the pACG_XylSPm plasmid series reported unsuccessful induction behavior, the pACG_araBAD plasmid series were selected for expression regulation. Plasmid pSW3_ *lacI*⁺ expressing recombinant mini-proinsulin was additionally transformed into the aforementioned mutants in order to evaluate effect of genetic regulation on ncBCAA mis-incorporation. At that stage a total of seven *geneX*-tunable *E. coli* strains were engineered, one for each target gene. Each engineered tunable *E. coli* strain lacked a certain endogenous target gene, generally indicated as *geneX* in this work (*E. coli* K-12 BW25113 $\Delta geneX$), and contained one plasmid allowing exogenous expression regulation of such target gene (pACG_araBAD_ *geneX*) and one plasmid enabling expression of recombinant mini-proinsulin for subsequent analysis of the impurity profile (pSW3_ *lacI*⁺). The engineered tunable *E. coli* mutants were screened in a mini-reactor by triggering induction of different expression levels of the target genes. The impurity profile of recombinant mini-proinsulin was then compared with the control non-engineered *E. coli* host. Screening was performed in fed-batch mode under standard cultivation conditions and under conditions mimicking large-scale effects i. e. pyruvate pulsing and O₂ limitation. After screening in the mini-reactor system, potential tunable *E. coli* strains expressing a certain level of the target gene resulting in an improved impurity profile were scaled-up in a 15L reactor and impurity profile of the recombinant protein was analysed in order to confirm reproducible behaviour. Validated *E. coli* strains reporting a reduced ncBCAA mis-incorporation may have a crucial role in the pharmaceutical industry since product quality is pivotal for commercial approval of recombinant proteins that are to be used as human therapeutics (Apostol *et al.*, 1997).

2. Literature Review

2.1 *E. coli* as industrial biofactory

2.1.1 Advantages of *E. coli* as industrial biofactory

The Gram-negative bacterium *E. coli* has been the most extensively used microbial host for recombinant protein manufacturing in the field of research and pharmaceutical industry, as well as for metabolic engineering purposes. Somatostatin was the first human recombinant protein to be expressed in this platform in 1977 (Itakura *et al.*, 1977) while humulin was the first commercially approved biopharmaceutical produced in *E. coli* in 1982 (Johnson, 1983). Since then, a total of 45 recombinant drugs produced in *E. coli* have been commercially approved for human use, representing about 29.8% of the market of recombinant biopharmaceuticals (Ferrer-Miralles, 2009). Lantus (Insulin glargin), also manufactured in *E. coli*, was the 4th top-selling biopharmaceutical product in 2013, with 7.95 billion \$ sales (Walsh, 2014), thus revealing the high impact of *E. coli* in the pharmaceutical industry, even nowadays, when the use of non-prokaryotic expression systems is increasing. There are several advantages making *E. coli* appealing as a host organism for protein production purposes. *E. coli* has been for years extensively characterized in relation to physiology and genetics, being its genomic sequence available and annotated (Blattner *et al.*, 1997). Its manipulation by recombinant DNA technology is greatly developed: various molecular techniques are available for genome editing, hence allowing strain improvement and reverse genetics (Madyagol *et al.*, 2011; Nakashima and Miyazaki, 2014), and easy protocols for transformation of *E. coli* with exogenous plasmid DNA are also well-established, thus facilitating expression of a wide range of recombinant proteins (Cohen *et al.*, 1972; Hanahan, 1983; Fiedler and Wirth, 1988; Shiloach and Fass, 2005). *E. coli* features fast growth kinetics, with a minimal doubling time as short as 20 min (Sezonov *et al.*, 2007). Moreover, there is a wide range of cultivation procedures with inexpensive media enabling high cell density cultures leading to high productivity (Riesenbergs, 1991; Lee, 1996; Choi *et al.*, 2006). Furthermore, scalability of *E. coli* based bioprocesses up to industrial scale is easily achieved.

2.1.2 Origin and genetic background of strain *E. coli* K-12 BW25113

E. coli K-12 BW25113 is a frequently used laboratory strain, which was first reported by Datsenko and Wanner (2000). It is the parent strain for the so-called Keio Collection, which contains around 4000 single gene knock-out mutants (Baba *et al.*, 2006). The complete 4631469-bp genome sequence of strain *E. coli* K-12 BW25113 was made available by Grenier *et al.* (2014) and it was deposited in GenBank under accession number CP009273. Strain BW25113 was derived from strain BD792 after a 13-step process consisting of serial transduction and allele replacement steps. Strain BD792 was, in turn, derived from the ancestral *E. coli* K-12 (EMG2) after a 2-step process. Strain BD792 and its predecessors contain the *rpoS396(Am)* allele, which has a base pair substitution that causes an amber mutation, hence triggering premature translation termination of the stationary-phase sigma factor σ^S . However, strain BW25113 carries the pseudo-revertant *rpoS(Q33)* allele (Hayashi *et al.*, 2006). In a revertant mutant the original mutation itself is mutated back to the wild type version

while in a pseudo-revertant mutant the original mutation stays invariable while another mutation occurring in a different position within the same gene restores the wild type phenotype.

Strain *E. coli* K-12 BW25113 is characterized by the following genotype: *F*, λ^- , *lacI*⁺, $\Delta(araD-araB)567$, $\Delta(rhaD-rhaB)568$, $\Delta lacZ4787(::rrnB-3)$, *rph-1*, *ilvG-1*, *hsdR514*. It lacks the F-plasmid necessary for gene transfer and the genes for motility (*F*⁻) as well as the lambda lysogen (λ^-). This strain was originally reported to be *lacI*^q (Datsenko and Wanner, 2000), but it was later demonstrated to be *lacI*⁺ (Grenier *et al.* 2014; Baba *et al.*, 2006). *araBAD* and *rhaBAD* genetic regions were deleted while a section of *lacZ* was replaced with four tandem copies of the *rrnB* transcriptional terminator (Datsenko and Wanner, 2000). Deletion of *araBAD* extends from around 25 bp upstream of the *araB* start codon to about 8 bp into the beginning of the *araD* gene. Hence, the strain is defective in arabinose ($\Delta araBAD$), rhamnose ($\Delta rhaBAD$) and lactose ($\Delta lacZ$) metabolizing enzymes, which enables proper functionality of arabinose, rhamnose and lactose inducible promoters, respectively. Moreover, frameshift mutations in *rph*, *ilvG* and *hsdR* resulting in a premature translation stop codon were reported.

The *rph* gene encodes for an exoribonuclease which plays a crucial role in tRNA maturation (Bowden *et al.*, 2017) and rRNA degradation (Basturea *et al.*, 2011). *rph* is organized together with *pyrE* in the same operon. *pyrE* encodes for an orotate phosphoribosyltransferase, which is involved in the biosynthesis of pyrimidine nucleotides. An intercistronic *pyrE* attenuator is located downstream of *rph* before open reading frame of *pyrE* (Poulsen *et al.*, 1983). In *E. coli* K-12 strains, a GC base pair deletion close to the end of *rph* genetic region results in a premature translation termination. As a consequence, *rph* gene product is truncated (*rph*⁻) and lacks exoribonuclease activity. Moreover, the premature translation termination of *rph* leads to a deficiency in *pyrE* expression, since a proper coupling between *rph* transcription and translation is necessary in order to maintain optimal transcription levels of *pyrE*. *E. coli* K-12 strains are characterized by pyrimidine starvation when cultivated in medium lacking exogenous pyrimidine sources such as uracil (Jensen, 1993).

The *ilvG* gene encodes for acetohydroxy acid synthase II (AHAS II), which is involved in the biosynthesis of isoleucine starting from α -ketobutyrate. In *E. coli* K-12 strains, a two-base insertion event is present in the coding sequence of gene *ilvG* between base pair positions 1250 and 1253, resulting in a frameshift mutation. This insertion causes a shift of the reading frame and, as a consequence, a stop codon is formed, resulting in a premature termination of *ilvG* gene translation (*ilvG*⁻). AHAS II is hence not expressed and distal expression of *ilvEDA* operon is impaired. The absence of AHAS II activity in *E. coli* K-12 strains leads to the valine toxicity phenomenon: a valine-mediated inhibition of *E. coli* growth (Yoon *et al.*, 2012; Parekh and Hatfield, 1997; Lawther *et al.* 1981). However, frameshift mutations affecting *rph* and *ilvG* can be abolished by applying genome engineering strategies (Hirokawa *et al.*, 2013; Biryukova *et al.*, 2010). Genome sequencing and annotation studies revealed that, as opposed to *E. coli* K-12 variants, strain *E. coli* BL21(DE3) contains intact *ilvG* and *rph* genes. As a consequence, B strains grow better than K-12 variants in minimal media and growth is not inhibited by valine (Yoon *et al.*, 2012).

2.1.3 Inclusion body-based recombinant protein expression in *E. coli*

Recombinant proteins expressed in *E. coli* are generally accumulated in so-called inclusion bodies, which are highly stable insoluble protein aggregates resistant to proteolysis. Size of IB particles

ranges from 50 to 800 nm. IBs can be seen as electrodense dark particles polarly distributed in the cell under transmission electron microscope but they are often even visible in light microscopy (Rinas *et al.*, 2017). IB formation occurs due to disequilibrium between protein translation, folding and aggregation rates (Kiefhaber *et al.*, 1991). Protein aggregation mainly happens when folding intermediates displaying hydrophobic residues at their surface accumulate due to insufficient presence of chaperones. These folding intermediates are prone to aggregate with each other through strong intermolecular hydrophobic interactions, which confer high stability to IBs (Mayer and Buchner, 2004). It was generally considered that IBs comprise non-active protein forms. Nevertheless, a vast palette of publications reporting biologically active IBs was recently made available (reviewed by García-Fruitós, 2010). These evidences might benefit the enzymatic industry since active IBs containing recombinant enzymes could be directly employed as catalyzers without the necessity of tedious refolding steps (Martínez-Alonso *et al.*, 2009). IBs mainly contain the recombinant protein, which can represent up to 95 % of the total protein, but also minor cellular protein rests, phospholipids and nucleic acids, which co-precipitate during the IB isolation process. Among cellular proteins, membrane proteins are mostly present in the IB fraction, such as proteases OmpT, OmpF, OmpC and OmpA. In addition, cytoplasmic proteins can also be found in the IB fraction in lower levels. This is the case of the ribosomal subunit proteins L7/L12, the chaperons DnaK and GroEL as well as the heat-shock proteins IbpA and IbpB. Plasmid-encoded proteins involved in the acquirement of antibiotic resistance, such as β -lactamase precursor and kanamycin resistance protein can also be present in the IB fraction (Jürgen *et al.*, 2010; Rinas and Bailey, 1992; Hartley and Kane, 1988). Furthermore, the protein composition of IBs is dependent on a number of factors such as the selected recombinant protein to be expressed, the bacterial host, the cultivation conditions and the employed IB purification methods (Rinas *et al.*, 2017).

Although *E. coli* does not have an efficient secretion system to produce recombinant proteins extracellularly and despite that downstream processing of recombinant proteins present in IBs is complex, *E. coli* is still the first choice when expressing simple proteins which do not require significant post-translational modifications. The main advantage of IB-based recombinant protein production is that product is resistant to proteolytic degradation and it is obtained at high concentrations and almost pure. Moreover, physicochemical properties inherent to IBs enable an easy isolation from the cellular rests by using standard protocols. Additionally, IBs can also be used as a model for the study of conformational diseases, as drug delivery systems, biomaterials and immobilized catalysts (Rinas *et al.*, 2017; Ramón *et al.*, 2014). Novel approaches improving recovery of functional proteins from IBs are reviewed by Ramón *et al.* (2014) and are mainly based in the optimization of solubilization and refolding steps. Recent methodologies leading to a reduction of IB formation during recombinant protein production processes are summarized by Fahnert *et al.* (2004) and Zhu *et al.* (2013), including reduction of translation rate, use of fusion proteins, co-expression of chaperones and foldases, use of folding promoting agents, oxidizing cytoplasmic redox potential and alteration of cultivation conditions. Even strategies triggering IB formation, mainly consisting in the fusion of aggregation-prone peptides to the recombinant protein, are reviewed by Rinas *et al.* (2017).

2.1.4 Insulin expression in *E. coli*

Insuline is a peptide hormone which regulates the level of sugar in the blood. It stimulates glucose uptake by the adipose, muscle and liver tissues, thus having a crucial role in regulating metabolism of

carbohydrates, fats and proteins. The pre-pro-insulin precursor of human insulin (PPI) is encoded by gene *INS*. It is synthesized in the endoplasmic reticulum (ER) of beta cells present in the islets of Langerhans located in the pancreas. PPI is a biologically inactive form of insulin, containing 110 amino acids organized in A-chain, C-peptide, B-chain and an N-terminal signal peptide, which guides the nascent protein chain into the ER lumen. The signal peptide is then cleaved by signal peptidases, yielding pro-insulin (PI). Conditions present in the ER lumen enable folding of PI into the proper conformation as well as formation of disulfide bonds. PI is then transported through the Golgi apparatus, where maturation of PI into active insulin takes place. Various peptidases trigger the cleavage of the C-peptide, which is located between A- and B-chains, yielding insulin. The C-peptide has a crucial role in the insulin biosynthesis since it joins both A- and B-chains in a manner that allows proper folding and interchain interaction through disulfide bond formation. The C-peptide itself has also been reported to activate Ca^{2+} -dependent intracellular signaling pathways upon binding to cell surface receptors as well as to enhance the Na^+ - K^+ -ATPase and the endothelial nitric oxide synthase activities, among others (Wahren *et al.*, 2000). The active insulin form is then transported by secretory vesicles through the cell, being finally secreted into the blood circulation. The human insulin has 51 amino acids, comprising A- and B-polypeptide chains, which remain linked by 2 disulfide bonds. A-chain also contains an intrapeptide disulfide bond (Beals *et al.*, 2008; Zündorf and Dingermann, 2001).

Since 1920s insulin was extracted from pork and beef pancreas glands collected from rests of the meat industry. However, insulin from animal sources could trigger immunologic sensitization when injected in human patients. Moreover, animal sources were not sufficient to supply an increasing diabetic population and it was then necessary to explore alternative sources of insulin. The successful expression of human insulin in *E. coli* by applying the recombinant DNA technology was reported in 1978, ensuring a reliable, safe and constant supply of insulin to diabetic patients around the globe (Johnson, 1983). Human insulin was the first commercially approved biopharmaceutical produced in *E. coli* (1982) and, since then, it has been one of the top-selling biopharmaceutical products on the market, which is in accordance with the increasing prevalence of diabetes among adults (Walsh, 2014).

When expressing recombinant human insulin in *E. coli*, a modified gene sequence different from the original one is used. This altered sequence is adapted to the *E. coli* codon usage and the genetic region encoding for the signal peptide is normally substituted by the methionine codon (ATG), which corresponds to the start codon for protein translation. Pro-insuline gene is then cloned into an expression plasmid under the control of a strong inducible promoter. Moreover, expression plasmids normally contain an antibiotic selection marker in order to ensure plasmid stability (Zündorf and Dingermann 2001). High yield expression of recombinant pro-insulin in *E. coli* leads to pro-insulin aggregation and intracellular IB formation due to the lack of chaperones and the high concentration of folding protein (Mayer and Buchner, 2004). Pro-insulin-containing IBs can represent as much as 20 % of the total *E. coli* cellular volume (Williams *et al.* 1982). Upon completion of *E. coli* cultivation, IBs have to be isolated from the *E. coli* cellular rests by cell disruption and then be solubilized in order to release the free pro-insulin molecules. However, at that stage, recombinant pro-insulin is not active since the complex mechanism leading to maturation of insulin precursors is not available in *E. coli*. Hence, steps conferring the functional peptide conformation taking place in the pancreatic beta cells have to be reproduced *ex vivo*: pro-insulin is refolded into the native conformation, the first methionine residue at N-terminal is removed by CNBr or CNCl treatment, the disulfide bond

formation is triggered by the so-called oxidative sulfitolysis and the C-peptide is cleaved by treatment with trypsin and carboxypeptidase B (Figure 1) (Mayer and Buchner, 2004; Vallejo and Rinas, 2004; Zündorf and Dingermann, 2001).

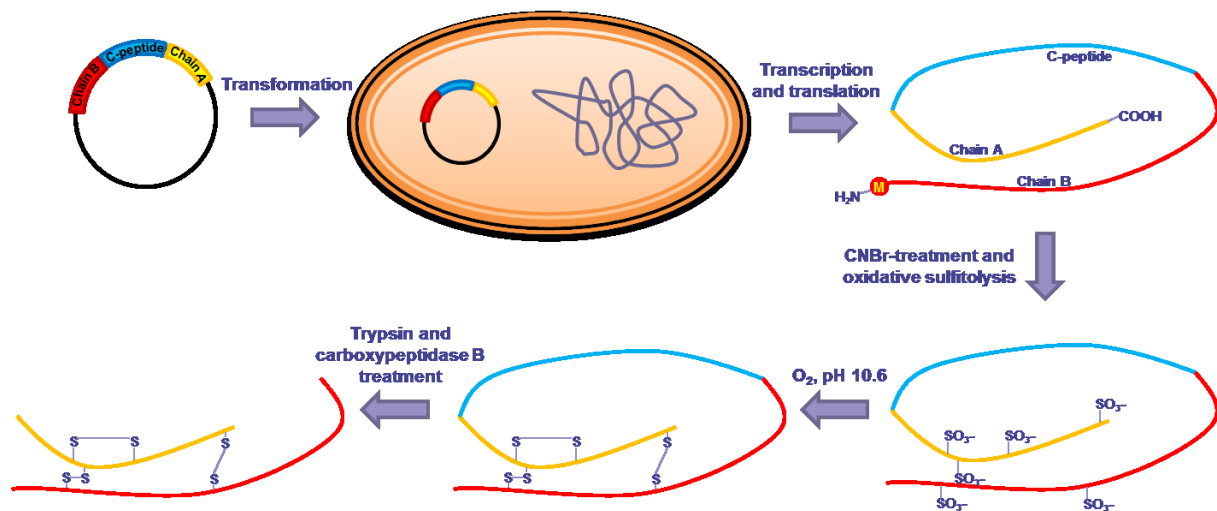


Figure 1. Expression process of recombinant insulin in *E. coli*. Adapted from Zündorf and Dingermann (2001).

2.1.5 Physiological stress during recombinant protein expression in *E. coli*

Recombinant protein production is an artificial stress imposed to the *E. coli* metabolism, causing metabolic burden, which refers to the drainage of cell resources (energy molecules such as ATP, GTP and NAD(P)H; and/or amino acid pools) towards the maintenance and expression of exogenous DNA within that cell. This metabolic imbalance triggers a vast number of effects: growth arrest due to reduction in the synthesis of biomass-related proteins, limited recombinant protein biosynthesis, plasmid instability, accumulation of by-products, ribosome destruction, protein hydrolysis and trigger of stress responses such as the ppGpp-mediated stringent response (Carneiro *et al.*, 2013). Main effects of the stringent response comprise inhibition of RNA polymerase and rRNA/tRNA synthesis, translation repression as well as activation of biosynthetic operons (Lengeler *et al.*, 1999; Traxler *et al.*, 2008). Metabolic burden reported in *E. coli*-based recombinant protein production processes might be similar to the naturally occurring bacteriophage infection, since phages capture the *E. coli* metabolic network in order to proliferate and spread infection (Fahnert *et al.* 2004). However, available engineering strategies leading to a reduction of the metabolic burden effects are reviewed by Carneiro *et al.* (2013).

Misfolded protein variants can accumulate in the *E. coli* cytosol either by protein aggregation (e.g. due to recombinant protein production processes) or by protein denaturation (e.g. due to high temperature). Both cases trigger the induction of the so-called heat shock response. This response comprises the stabilization and secondary self-induction of the heat-shock sigma factor (σ^{32}) activity, which, in turn, enhances transcription of numerous heat shock proteins, including Hsp70 and Hsp100 chaperones, Hsp60 chaperonins as well as ATP-dependent Lon and ClpAB proteases. Synthesized heat shock proteins enable either refolding or degradation of the misfolded protein forms (Fahnert *et al.* 2004; Mogk *et al.*, 2015). Likewise, accumulation of misfolded protein variants in the periplasm induces σ^{24} -dependent envelope shock response (Alba and Gross, 2004).

The general stress response depends on the sigma regulator σ^s , which activates transcription of genes inducing transition to stationary phase. General stress response is activated in non-recombinant *E. coli* cultivations during glucose starvation and in glucose-limited fed-batch cultivations (Teich *et al.* 1999). Moreover, Gill *et al.* (2000) reported a significant up-regulation of σ^s for various *E. coli* cultivations producing different recombinant proteins. Interestingly, Schweder *et al.* (2002) showed that the general stress response is not induced during very strong expression of recombinant α -glucosidase.

Moreover, Arís *et al.* (1998) demonstrated that recombinant protein production driven by strong lambda lytic promoters can promote activation of the SOS response. The SOS proteins trigger DNA repair and resume cell division once DNA replication is properly restored. Nevertheless, when metabolic burden associated to recombinant protein expression exceeds a limit, cells are not able to induce the SOS response, even if DNA damage occurs (Lin *et al.*, 2001).

2.2 Introduction to central *E. coli* metabolism

2.2.1 Catabolite repression and oxidative respiration

E. coli can metabolize a number of alternative carbon sources in order to obtain energy and synthesize endogenous constituents for cell growth and maintenance. However, *E. coli* prefers glucose than any other carbon source so that, when glucose and other carbon sources are simultaneously present in the media, *E. coli* first utilizes glucose and then, the alternative carbon sources. Glucose preference relies on the catabolite repression mechanism, which ensures expression inhibition of catabolic operons of carbon sources different than the preferred one. Glucose depletion (i.e. starvation or limitation) causes de-repression of the catabolic operon of the next preferred carbon source, hence allowing its utilization.

The catabolite repression mechanism is mainly regulated by 3 enzymes involved in the PEP-dependent carbohydrate phosphotransferase system (PTS), including sugar-specific PTS permease or enzyme II (EII), enzyme I (EI) and histidine-containing protein (HPr). These enzymes sequentially catalyze the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) to the imported sugar, yielding pyruvate and sugar phosphates. EII comprises 3 protein domains (EIIA, B and C). The phosphorylated form of EIIA regulates activity of adenylate cyclase, which catalyzes the transformation of ATP to cAMP and pyrophosphate. In *E. coli*, catabolite activator protein (CAP) is a global regulator protein activating the transcription of a variety of genes. cAMP acts as an allosteric effector and, when bound to CAP, increases CAP's affinity for DNA, hence allowing CAP's interaction with the CAP-binding site, located in the promoter region. CAP can then activate transcription by interacting with RNA polymerase (Brückner and Titgemeyer, 2002).

The catabolite repression mechanism affecting the lactose operon is presented as the classical example in *E. coli*. On the one hand, when glucose is present in the medium, EIIA remains in unphosphorylated form and adenylate cyclase is inhibited, cAMP levels remain low and CAP does not bind to the promoter of the lactose operon, hence avoiding lactose consumption. On the other hand, when glucose is depleted, EIIA switches to the phosphorylated form and adenylate cyclase is activated, cAMP levels increase and CAP activates transcription of lactose catabolic enzymes.

Glycolysis is the principal pathway for glucose catabolism. It is an oxygen-independent pathway comprising 10 sequential enzymatic reactions catalyzing conversion of glucose into pyruvate and release of ATP and NADH molecules. Depending on the environmental conditions, pyruvate can be further metabolized through two main catabolic pathways: oxidative respiration or fermentation. In the presence of oxygen, oxidative respiration is activated and pyruvate is decarboxylated to acetyl-CoA by the pyruvate dehydrogenase complex (PDHC, encoded by genes *aceE*, *aceF* and *lpdA*), generating CO₂ as by-product. Expression of PDHC is down-regulated in the absence of oxygen while its activity is enhanced by pyruvate (Quail, Haydon, and Guest 1994). Acetyl-CoA can enter the Krebs cycle, which begins with acetyl-CoA reacting with oxaloacetate to yield citrate and CoA. This reaction is catalyzed by citrate synthase (CISY, encoded by gene *gltA*). The Krebs cycle is an 8-step process catalyzing oxidation of acetyl-CoA to CO₂ and release of ATP and NADH. During glycolysis and Krebs cycle numerous intermediates are formed, which serve as precursors for anabolic pathways. The generated NADH is then directed to the oxidative phosphorylation pathway. Electrons are transferred from NADH to oxygen through an electron transport chain, generating ATP. Oxidation of NADH to NAD⁺ releases protons into the cytoplasm and energy generated by the electron transport chain pumps protons across the membrane into the periplasmic space, generating a transmembrane electrochemical gradient. Protons flow then back across the membrane through ATP synthase, resulting in ATP generation (Madigan *et al.* 2014).

2.2.2 Mixed-acid fermentation

In the absence of oxygen as electron acceptor, mixed-acid fermentation is the metabolic pathway employed for ATP generation in *E. coli*. However, when it comes to energy production, oxidative respiration is preferred to mixed-acid fermentation since the first is much more efficient (Sawers *et al.*, 2004). Mixed-acid fermentation involves the transformation of a hexose into a mix of end fermentation products, such as acetate, lactate, formate, succinate, CO₂ and H₂ (Figure 2). All mentioned fermentation products derive from pyruvate either directly or via acetyl-CoA, with exception of succinate, which originates from phosphoenolpyruvate via oxaloacetate (Xu *et al.* 1999). Each generated fermentation product requires a different set of enzymes to be produced. Fermentation products are finally secreted extracellularly. The proportion of each end product relies on a variety of factors, including relative activity of enzymes involved in the fermentation process, availability of electron acceptors, oxidation form of substrates and presence of redox agents (Liu *et al.* 2011). Moreover, bacterial strain can also affect such relative proportion. For instance, acetate is the dominant product in *E. coli* K-12 strains and it has been considered one of the major inconvenients in *E. coli*-based recombinant protein production processes due to its toxicity (Phue *et al.* 2005; Wolfe, 2005).

Under anaerobic conditions, and especially at low pH, lactate dehydrogenase (LDH), encoded by gene *ldhA*, catalyzes the conversion of pyruvate into lactate while oxidizing NADH into NAD⁺, hence recycling the NADH generated during glycolysis. Formate production is accomplished by pyruvate-formate lyase (PFL), encoded by gene *pflB*. PFL catalyzes the transfer of the acetyl group of pyruvate to CoA yielding formate and acetyl-CoA. Enzyme activity of LDH and PFL is inhibited by the presence of oxygen and repressed by feedback regulation (Kessler and Knappe 1996). The expression of gene *pflB* is regulated by pyruvate accumulation under oxygen limitation conditions (Sirko *et al.* 1993). PFL expression is reported as the most sensitive response when modifying oxygen availability in *E. coli*.

based cultivations (Schweder *et al.* 1999). Formate can be further metabolized into CO₂ and H₂ by the formate hydrogenlyase complex (FHL), whose functionality strongly relies on the adequate content of trace elements molybdenum, nickel and selenium in the cultivation medium (Biermann *et al.*, 2013). FHL complex is a multimeric protein consisting of numerous subunits encoded by genes *fdhF* and *hycB-G*. Mnatsakanyan *et al.* (2004) showed that FHL plays a key role in H₂ and CO₂ production at acidic pH. Acetate synthesis is assisted by the sequential action of two enzymes: phosphate acetyltransferase (encoded by gene *pta*) and acetate kinase (encoded by gene *ackA*). First, phosphate acetyltransferase reversibly converts acetyl-CoA and phosphate into acetyl-phosphate and CoA, respectively. Second, acetate kinase reversibly transforms acetyl-phosphate into acetate, while generating ATP. Pyruvate and PEP play as activators of phosphate acetyltransferase (Campos-Bermudez *et al.*, 2010). Among all mixed-acid fermentation reactions, the acetate producing pathway is the only one generating energy in form of ATP. Hence, under anaerobic growth conditions, acetate kinase is involved in the synthesis of most of the ATP obtained by catabolism. Ethanol formation is catalyzed by alcohol dehydrogenase (ADH), encoded by gene *adhE*, in a 2-step reaction. In the first step acetyl-CoA is converted to acetaldehyde and CoA while oxidizing NADH into NAD⁺. In the second step acetaldehyde is converted to ethanol while oxidizing NADH into NAD⁺. Succinate synthesis involves action of four different enzymes. First, phosphoenolpyruvate (PEP), a glycolysis intermediate, is carboxylated by PEP carboxylase (encoded by gene *ppc*) to yield oxaloacetate. Second, oxaloacetate is transformed to malate by malate dehydrogenase (encoded by gene *mdh*), while oxidizing NADH into NAD⁺. Third, fumarate hydratase (encoded by gene *fumB*) catalyzes dehydration of malate to form fumarate. Finally, fumarate is converted to succinate thanks to fumarate reductase complex (encoded by genes *frdA-D*), while oxidizing NADH into NAD⁺.

2.2.3 Overflow metabolism

Overflow metabolism refers to the phenomenon where incomplete glucose oxidation by fermentation is favored instead of the more energetically efficient complete oxidation by oxidative respiration, even under aerobic conditions. When oxygen is present but glucose is in excess in the cultivation medium, glycolysis and PDHC reactions are stimulated, resulting in high amounts of ATP and NADH as well as accumulation of pyruvate and acetyl-CoA. This decreases the need for production of additional energy and reducing equivalents by the Krebs cycle. As a consequence, the Krebs cycle is blocked and oxygen consumption is reduced. Under this scenario, NAD⁺ and CoA pools are significantly drained. However, the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH, encoded by gene *gapA*) requires NAD⁺ and PDHC requires CoA as substrate, respectively. Hence, reoxidation of NADH into NAD⁺ and recycling of CoA must be ensured in order to maintain the glycolytic flux towards pyruvate and acetyl-CoA. Moreover, excess of pyruvate and acetyl-CoA must be consumed. Since Krebs cycle is inhibited, this is achieved by activating the reactions of the mixed-acid fermentation (Figure 2). While acetate production serves to generate ATP, formation of other mixed-acid fermentation products regenerates NAD⁺ and CoA pools (Wolfe, 2005).

In *E. coli* acetate is the predominant product of the overflow metabolism (B. Xu *et al.* 1999) and it is mainly generated by the sequential action of phosphate acetyltransferase and acetate kinase in the so-called PTA-ACKA pathway. Alternatively, pyruvate oxidase (encoded by gene *poxB*) triggers the direct oxidative decarboxylation of pyruvate into acetate and CO₂ under aerobic conditions, while

reducing FAD to FADH₂. The reversible PTA-ACKA pathway can also catabolize acetate, but much less efficiently than the Acetyl-CoA synthetase pathway (Brown *et al.*, 1977). Acetyl-CoA synthetase (ACS, encoded by gene *acs*) converts acetate in acetyl-CoA by consuming ATP and it is subjected to catabolite repression (Valgepea *et al.*, 2010). Overflow metabolism causes an intracellular accumulation of pyruvate which, in turn, activates PDHC and PTA (Quail *et al.*, 1994; Campos-Bermudez *et al.*, 2010), resulting in a significant acetate production. *E. coli* cells switch from acetate production (overflow metabolism) to acetate consumption modus mainly by regulating expression of *acs*. Activity of ACS is repressed when glucose is in excess while it is induced when glucose is depleted and acetate is present. Hence, ACS enables the cell to use acetate under aerobic conditions in order to generate energy via the Krebs cycle, once glucose is depleted (Wolfe, 2005).

Acetate accumulation has proven to be toxic for *E. coli* cells, having negative consequences at physiological level. Acetate can easily cross cell membrane, altering the transmembrane pH gradient. Moreover, it causes cytoplasm acidification, increases osmotic pressure and jeopardizes methionine biosynthesis (reviewed by Wolfe, 2005). In *E. coli*-based recombinant protein production processes acetate accumulation results in a reduction of growth rate and product yields (Eiteman and Altman, 2006).

Furthermore, the intensity of overflow metabolism and the subsequent acetate accumulation is strain-dependent. *E. coli* K-12 strains are reported to have a higher acetate accumulation than B strains. This is commonly explained by three metabolic activities reported to be more active in B strains: acetate uptake by acetyl-CoA synthetase, glyoxylate shunt encoded by the acetate operon *aceBAK* and Krebs cycle (Phue *et al.* 2005; Yoon *et al.*, 2012).

Several strategies have been made available, leading to a reduction of overflow metabolism and acetate overproduction in *E. coli*, including process or strain engineering approaches (reviewed by Eiteman and Altman, 2006). Recently, Anane *et al.* (2017) elaborated a model describing overflow metabolism.

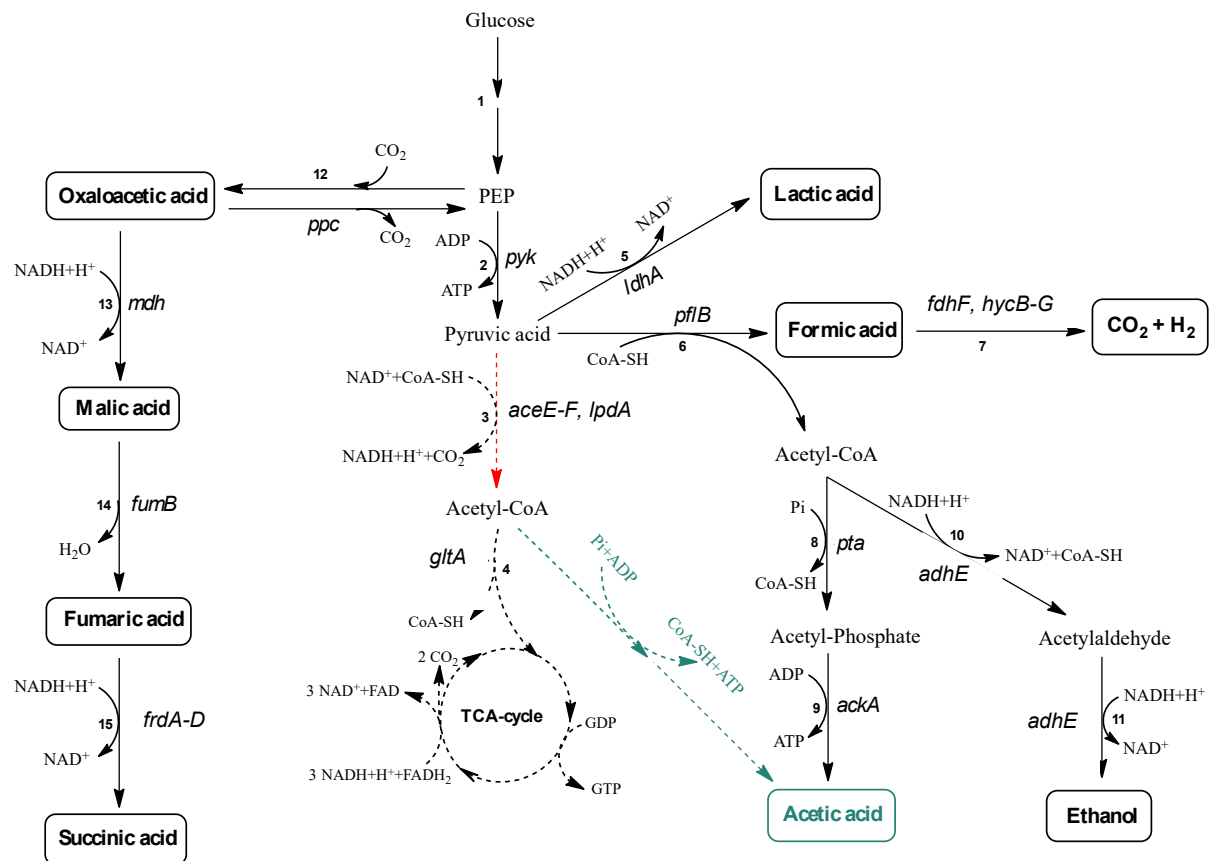


Figure 2. Schematic representation of mixed-acid fermentation and overflow metabolism pathways present in *E. coli*. Dashed red and black lines display pyruvate flux towards Krebs cycle. Dashed green lines indicate acetate synthesis via overflow metabolism. 1: glycolytic enzymes; 2: pyruvate kinase; 3: pyruvate dehydrogenase complex; 4: citrate synthase; 5: lactate dehydrogenase; 6: pyruvate-formate lyase; 7: formate hydrogenlyase complex; 8: phosphate acetyltransferase; 9: acetate kinase; 10 and 11: alcohol dehydrogenase; 12: phosphoenolpyruvate carboxylase; 13: malate dehydrogenase; 14: fumarate hydratase; 15: fumarate reductase complex. Genes encoding aforementioned enzymes appear in italics. Adapted from Reitz (2017).

2.2.4 BCAA biosynthetic pathway

A schematic overview of the BCAA biosynthetic pathway is shown in Figure 3.

2.2.4.1 Biosynthesis of isoleucine and valine

The first step of the isoleucine biosynthesis pathway is catalyzed by threonine deaminase (encoded by gene *ilvA*), a pyridoxal 5'-phosphate-dependent enzyme which transforms L-threonine to α -ketobutyrate and ammonia in a two-step reaction. First, L-threonine is dehydrated to yield enamine/imine intermediates. Second, intermediates are deaminated to form α -ketobutyrate and ammonia (Umbarger and Brown, 1957). An alternative source of α -ketobutyrate is the direct carbon chain elongation of pyruvate catalyzed by enzymes encoded by the *leuABCD* operon, which comprises three enzymatic reactions. First, α -isopropylmalate synthase (encoded by gene *leuA*) condensates the acetyl group of acetyl-CoA with pyruvate to form citramalate and CoA. Second, α -isopropylmalate isomerase (encoded by genes *leuC-D*) catalyzes a 2-step isomerization of citramalate to β -methyl-D-malate, via citraconate. Third, a NAD-dependent β -isopropylmalate dehydrogenase

(encoded by gene *leuB*) catalyzes the oxidative decarboxylation of β -methyl-D-malate to produce α -ketobutyrate (Bogosian *et al.* 1989).

After formation of α -ketobutyrate, the following four reactions of the isoleucine biosynthesis pathway are driven by the same enzymes catalyzing the parallel steps of the valine biosynthesis pathway.

First, acetohydroxy acid synthase (encoded by operons *ilvBN*, *ilvGM* and *ilvIH*) catalyzes a 2-step bi-substrate enzymatic reaction. In the first reaction step pyruvate binds to AHAS, undergoing decarboxylation to form an active intermediate. In the second reaction step two alternative substrates can bind to the pre-formed intermediate to generate two alternative acetohydroxy acids: a second pyruvate molecule to produce α -acetolactate as end-product or α -ketobutyrate to generate α -acetohydroxybutyrate. α -acetolactate serves as precursor for valine biosynthesis while α -acetohydroxybutyrate is a precursor of isoleucine. *E. coli* encodes three AHAS isozymes which diverge in their biochemical properties. Enzymes AHAS I, II and III are encoded by operons *ilvBN*, *ilvGM* and *ilvIH*, respectively. AHAS I has substrate preference for pyruvate in the second reaction step while AHAS II and III have a substrate preference for α -ketobutyrate, being this higher for AHAS II. Hence, AHAS I directs the metabolic flux to the production of α -acetolactate while AHAS II and III preferably to the formation of α -acetohydroxybutyrate. In addition, catalytic efficiency (given by the k_{cat}/K_m ratio) of AHAS I is 2-fold and 40-fold higher than AHAS II and AHAS III, respectively (Barak *et al.*, 1987; Salmon *et al.*, 2006; Vinogradov *et al.*, 2006).

Second, acetohydroxy acid isomeroreductase (encoded by gene *ilvC*) catalyzes the transfer of the ethyl group of α -acetohydroxybutyrate or the methyl group of α -acetolactate from the α -carbon to the β -carbon to form the α,β -dihydroxy acids α,β -dihydroxy- β -methylvalerate and α,β -dihydroxyisovalerate, respectively.

Third, dihydroxyacid dehydratase (encoded by gene *ilvD*) catalyzes dehydration of the α,β -dihydroxy acid intermediates to yield α -keto acids α -ketoisovalerate (valine precursor) and α,β -keto- β -methylvalerate (isoleucine precursor).

Fourth, the branched-chain amino acid aminotransferase or transaminase B (encoded by gene *ilvE*) catalyzes transamination of α,β -dihydroxy acids by using glutamate as donor of the amino group. Hence, α -ketoisovalerate is converted into valine while α,β -keto- β -methylvalerate is transformed into isoleucine. Moreover, valine can also undergo transamination by transaminase C (encoded by gene *avtA*) by using alanine as donor of the amino group (Whalen *et al.*, 1982).

2.2.4.2 Biosynthesis of leucine

Leucine biosynthesis comprises the carbon chain elongation of α -ketoisovalerate is catalyzed by enzymes encoded by the *leuABCD* operon, which comprises three enzymatic reactions. First, α -isopropylmalate synthase condensates the acetyl group of acetyl-CoA with α -ketoisovalerate to yield α -isopropylmalate and CoA. Second, α -isopropylmalate isomerase catalyzes a 2-step isomerization of α -isopropylmalate to β -isopropylmalate, via dimethyl-citraconate. Third, the NAD-dependent β -isopropylmalate dehydrogenase catalyzes the oxidative decarboxylation of β -isopropylmalate to produce α -ketoisocaproate. Transaminase B can then catalyze transamination of α -ketoisocaproate by using glutamate as donor of the amino group. Hence, α -ketoisocaproate is converted into leucine.

Furthermore, α -ketoisocaproate can also be subjected to transamination by the aromatic amino acid aminotransferase (encoded by gene *tyrB*) by using also glutamate as donor of the amino group (Vartak *et al.*, 1991; Huang *et al.*, 2009).

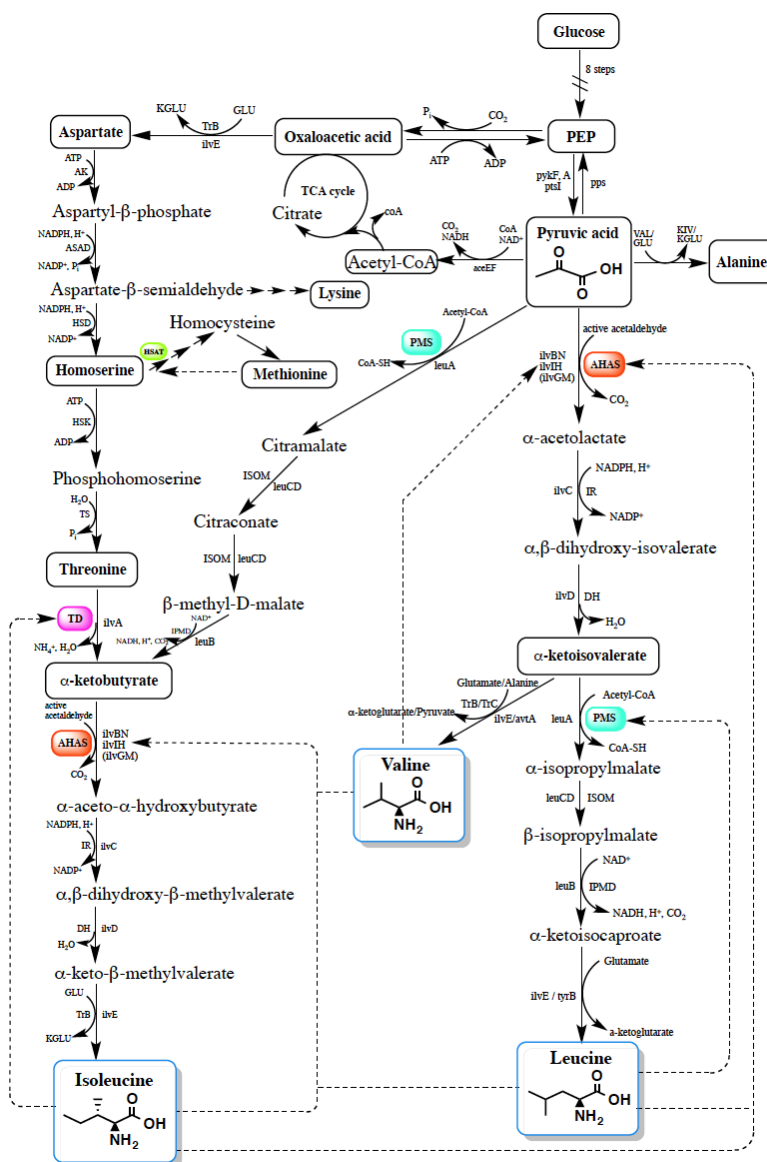


Figure 3. Schematic representation of the BCAA biosynthetic pathway. See text for description. Dashed lines indicated BCAA-mediated feedback regulation. AHAS: acetohydroxy acid synthase, encoded by genes *ilvBN*, *ilvGM* and *ilvIH*; DH: dihydroxy-acid dehydratase, encoded by *ilvD*; PMS: α -isopropylmalate synthase, encoded by *leuA*; IR: ketol-acid reductoisomerase (NADP(+)), encoded by *ilvC*; IPMD: 3-isopropylmalate dehydrogenase; encoded by *leuB*; ISOM: 3-isopropylmalate dehydratase, encoded by *leuCD*; TD: threonine deaminase, encoded by *ilvA*; TrB: transaminase B, encoded by *ilvE*; AK: bifunctional aspartokinase/homoserine dehydrogenase 1, encoded by *thrA*; ASAD: aspartate-semialdehyde dehydrogenase, encoded by gene *asd*; HSK: homoserine kinase, encoded by *thrB*; TS: threonine synthase, encoded by *thrC*; TrC: transaminase C, encoded by *avt*; *tyrB*: gene encoding aromatic amino acid aminotransferase. Adapted from Reitz (2017).

2.2.5 Regulation of the BCAA biosynthetic pathway

The branched chain amino acid biosynthesis pathway is effectively modulated by a complex regulatory network. This enables cells to adapt BCAA biosynthesis to shifting environmental conditions. Regulation can take place at enzymatic or genetic level. Genetic regulation is hierarchical and its range of action can be either global or affect local pathways. A summary of the main regulation mechanisms affecting the target genes investigated in this thesis is provided in Table 1.

Table 1. Summary of the regulation mechanisms affecting the target genes investigated in this thesis.

Gene	Enzyme	Transcriptional regulation	Posttranslational regulation
<i>leuA</i>	2-isopropylmalate synthase	-Attenuation (by leu) -Lrp regulation (leu does not show a modulation effect) -RelA/SpoT modulon (up-regulation by (p)ppGpp after amino acid starvation)	-Feed-back inhibition (by leu)
<i>ilvC</i>	Ketol-acid reductoisomerase (NADP(+))	-Substrate-mediated activation (substrate binding to a preformed IlvY protein DNA complex relaxes an IlvY protein-induced DNA bend and increases the affinity for RNA polymerase) -RelA/SpoT modulon (up-regulation by (p)ppGpp after amino acid starvation)	-
<i>ilvIH</i>	acetohydroxyacid synthase isozyme III	-Lrp regulation (leu inhibits Lrp activator from binding to the promoter regulatory region of <i>ilvIH</i> operon thus decreasing transcription) -RelA/SpoT modulon (down-regulation by (p)ppGpp after amino acid starvation)	-Feed-back inhibition (by val, leu and ile)
<i>ilvGM</i>	acetohydroxyacid synthase isozyme II	-Attenuation (val, leu and ile) -Lrp regulation (leu inhibits Lrp repressor from binding to the transcription site of <i>ilvGMEDA</i> operon thus increasing transcription) -IHF-mediated global regulation (activation) -RelA/SpoT modulon (up-regulation by (p)ppGpp after amino acid starvation)	-
<i>ilvBN</i>	acetohydroxyacid synthase isozyme I	-Attenuation (by val and leu) -CAP-mediated global regulation (activation)	-Feed-back inhibition (by val)

		when glc levels are low)	
		-IHF-mediated global regulation (activation)	
		-RelA/SpoT modulon (up-regulation by (p)ppGpp after amino acid starvation)	
<i>thrA</i>	threonine-sensitive, bifunctional aspartokinase/homoserine dehydrogenase 1	-Attenuation (thr and ile) -RelA/SpoT modulon (up-regulation by (p)ppGpp after amino acid starvation)	-Feed-back inhibition (thr). -Activation by ile and met.
<i>ilvA</i>	L-threonine dehydratase	- -	-Feed-back inhibition (ile). Activation by valine.

2.2.5.1 Posttranslational regulation

Allosteric regulation is an important mechanism of posttranslational regulation, which enables cells to instantaneously adjust their metabolism in response to environmental changes. The activity of the so-called allosteric enzymes can be adjusted by the non-covalent interaction of effector molecules to regulatory sites located in the enzyme outside the active site. Binding of the effector (activator or inhibitor) to the regulatory site triggers conformational changes which might either increase or decrease enzymatic activity. Allosteric regulation is typical of biosynthetic enzymes. In this case, the end-product of the biosynthetic pathway normally acts as inhibitor of the enzyme involved in that pathway. This regulation mechanism is known as feedback inhibition and it usually affects the first enzyme of a certain biosynthetic pathway. However, allosteric activation of a biosynthetic enzyme is also not rare (Lengeler *et al.*, 1999). Accordingly to aforementioned, various enzymes involved in the BCAA biosynthetic pathway are subjected to feedback inhibition and allosteric activation (Table 2).

Table 2. Enzymes involved in the BCAA biosynthetic pathway subjected to allosteric regulation and their corresponding inhibitors and activators.

Enzyme	Coding gene	Inhibitor	Activator	Reference
Aspartokinase/homoserine dehydrogenase	<i>thrA</i>	threonine	Isoleucine, methionine	Patte (1996)
Threonine deaminase	<i>ilvA</i>	isoleucine	valine	Umbarger (1956), Monod <i>et al.</i> (1965)
α -isopropylmalate synthase	<i>leuA</i>	leucine	-	Soper <i>et al.</i> (1976)
Acetohydroxy acid synthase	<i>ilvIH</i>	valine, leucine,	-	Salmon <i>et al.</i>

isozyme III	isoleucine			(2006)
Acetohydroxy acid synthase isozyme I	<i>ilvBN</i>	valine	-	Salmon <i>et al.</i> (2006)

2.2.5.2 Genetic regulation

In addition to posttranslational regulation, *e.g.* regulation of enzymatic activity, gene expression patterns are also strictly regulated in order to ensure adaptation of cells to nutritional and environmental changing conditions. As opposed to enzyme regulation, genetic regulation reports a lag time until the response mechanism to changes is effective but it also ensures a long-term response. Genetic regulation occurs at different hierarchical levels: local, regional and global. Local genetic regulation affects only a single gene or operon while regional regulation acts on multiple operons belonging to the same regulon. In an operon, genes coding for similar functions are organized as a single transcriptional unit (*e.g.* genes involved in lactose catabolism). A regulon comprises a set of operons which are dispersed in the genome but are under the control of the same specific regulator (*e.g.* operons involved in the biosynthesis of leucine). The integration of genes in operons and regulons enables an efficient and coordinated regulation of functionally related genes. Global genetic regulation affects the expression of modulones, which integrate genes, operons and/or regulons having a common global purpose (*e.g.* genes involved in anaerobic respiration) (Lengeler *et al.*, 1999). Genetic units belonging to a modulon are regulated by the same global regulator. Moreover, a stimulon is a group of genes responding to the same environmental stimulus (*e.g.* temperature).

A total of 15 genes involved in BCAA biosynthesis are grouped in 5 operons (*ilvBN*, *ilvGMEDA*, *ilvIH*, *ilvYC* and *leuABCD*), which, in turn, are integrated in the so-called *ilv* regulon. Transcription of operons *ilvBN*, *ilvGMEDA*, *leuABCD* and *thrABC* is regulated by attenuation (Vitreschak *et al.*, 2006). Operon *ilvYC* is regulated by an operon-specific mechanism mediated by protein IlvY. In addition, operons *ilvIH*, *ilvGMEDA* and *leuABCD* are regulated by the global Lrp (leucine-responsive protein). *ilvBN* operon is subjected to global catabolite repression control mediated by CAP (catabolite activator protein) as well. Moreover, the global IHF (integration host factor) also regulates expression of operons *ilvBN* and *ilvGMEDA* (Salmon *et al.*, 2006). Additionally, the global RelA/SpoT modulon is also described to regulate transcription of operons implicated in amino acid biosynthesis (Fang and Bauer, 2018).

2.2.5.2.1 Transcription attenuation

Transcription attenuation is a genetic regulation mechanism based in the synchronization of transcription and translation processes. The nascent RNA transcript of an operon contains a leader region located between the transcription initiation site and the first coding gene of the operon. The sequence of the leader region normally contains codons corresponding to amino acids synthesized by the gene products of the operon, i.e. regulatory codons. Shortly after transcription begins, a 1:2 stem-loop secondary structure is assembled, resulting in the obstruction of RNA polymerase to

proceed with the transcription process. RNA polymerase remains then momentarily paused at the end of stem-loop 1:2. This interruption enables the ribosome to start translating the leader peptide. When ribosome reaches steam 1, RNA polymerase resumes transcription and, from then on, both transcription and translation processes are coupled. Depending on the availability of certain amino-acylated tRNAs in the cell, two alternative scenarios can occur next. If the cellular level of necessary amino-acylated tRNAs is low, ribosome stops at regulatory codons while RNA polymerase continues transcribing, causing a desynchronization. This scenario allows formation of the anti-terminator 2:3 stem-loop secondary structure, hence impeding formation of transcription termination and enabling read-through into the structural genes of the operon. However, if the cellular level of necessary amino-acylated tRNAs is high, ribosome translates the regulatory codons and proceeds until the stop codon, thus translating the complete leader peptide. This scenario ensures synchronization of transcription and translation which avoids formation of the anti-terminator 2:3 stem-loop but triggers assembly of the terminator 3:4 stem-loop secondary structure, i.e. attenuator, resulting in a premature transcription interruption, i.e. attenuation (Vitreschak *et al.*, 2006; Salmon *et al.*, 2006; Lengeler *et al.*, 1999). Accordingly, ribosome is the main mediator of the transcription attenuation mechanism and the degree of attenuation relies on the position of the RNA polymerase when the ribosome dissociates from the RNA transcript and on the formation probabilities of each RNA secondary structure (Roesser *et al.*, 1989).

Transcription attenuation of an operon involved in the biosynthetic pathway of a certain amino acid (e.g. *leuABCD* operon) is usually mediated by the cellular levels of such amino acid, since the regulatory codons present in the leader region are codons for such amino acid. However, transcription attenuation of operons implicated in a biosynthetic pathway shared by various amino acids (e.g. *ilvBN* operon) might be mediated by all those amino acids, as the leader region would contain codons for such amino acids (Vitreschak *et al.*, 2006).

Transcription of operons *ilvBN*, *ilvGMEDA*, *leuABCD* and *thrABC* is regulated by attenuation. Table 3 shows the regulatory codons present in the leader peptide of the aforementioned operons.

Table 3. Regulatory codons present in the leader peptide of various operons involved in the BCAA biosynthetic pathway. Data obtained from Lengeler *et al.* (1999).

Operon	Regulatory codons	Amino acid sequence leader peptide
<i>ilvBN</i>	valine, leucine	MTTSM LNAK <u>LL</u> PTAPSAAV <u>VVV</u> R <u>VVVVV</u> GNAP
<i>ilvGMEDA</i>	valine, leucine, isoleucine	MT <u>ALL</u> R <u>V</u> IS <u>LV</u> VIS <u>VVV</u> VIIIPPCGAALGRGLA
<i>leuABCD</i>	leucine	MSHIVRFTG <u>LLL</u> NAFIVRGRPVGGI
<i>thrABC</i>	threonine, isoleucine	MKRIS <u>TTTTTT</u> ITTTGNGAG

2.2.5.2.2 *IlvY-mediated regulation of ilvYC operon*

As opposed to aforementioned operons, which are regulated by the end product of the specific biosynthetic pathway they are involved in, operon *ilvYC* is regulated by its substrates, e.g. α -

acetolactate. Genes *ilvY* and *ilvC* are organized in the same operon in opposite orientations and their transcription is controlled by the divergent overlapping *ilvYC* promoter. IlvY is a regulatory protein encoded by gene *ilvY* that can bind to operator sites O₁ and O₂, located in the promoter region. When no substrates of α -acetohydroxy acid isomeroreductase (encoded by *ilvC*) are present, ilvY strongly binds to the operator regions, forming a DNA bend that inactivates *ilvYC* promoter. When substrates are present, they can bind to the pre-formed IlvY-operator complex, resulting in a relaxation of the DNA bend and a dramatic increase of RNA polymerase affinity for *ilvC* promoter (Arfin *et al.*, 1969; Rhee *et al.*, 1998).

2.2.5.2.3 Global regulation

2.2.5.2.3.1 Lrp-mediated global regulation

The leucine-responsive regulatory protein (Lrp) is a global regulator of the transcription process. It regulates about 10 % of the *E. coli* genes, being most of them involved in the transport and metabolism of amino acids. Depending on the target operon, expression can be either activated or repressed upon Lrp action. Lrp recognizes and binds to DNA target sites generally located in the promoter region of the target operon. Leucine usually acts as an effector ligand of Lrp, being sometimes required for Lrp binding to target sites. However, in other cases, leucine can also impede or have no effect on the interaction of Lrp with the DNA target sites. In general, Lrp promotes expression of biosynthetic operons while repression of catabolic ones (Rhee *et al.*, 1996; de los Rios and Perona, 2006).

Operons *ilvIH*, *ilvGMEDA* and *leuABCD* were demonstrated to be regulated by the global Lrp protein in different manners (Lin *et al.*, 1992; Platko *et al.*, 1990; Rhee *et al.*, 1996). Lrp directly triggers activation of the *ilvIH* promoter and addition of leucine into the cultivation medium reduces binding capacity of Lrp to the target sites located at the *ilvIH* promoter region, resulting in a 5- to 10-fold transcription decrease (Chen *et al.*, 2005; DeFelice and Levinthal, 1977). As opposed to Lrp-mediated regulation of *ilvIH* operon, Rhee *et al.* (1996) showed that Lrp triggers transcriptional repression of the *ilvGMEDA* operon. Lrp binds to DNA in a position between the leader region and the first structural gene of the operon (*ilvG*). Moreover, leucine reduces binding capacity of Lrp to the target sites of the *ilvGMEDA* operon, resulting in a decrease of the repression degree. Lin *et al.* (1992) and Tchetina and Newman (1995) demonstrated that Lrp regulates expression of *leuABCD* operon. Different from aforementioned operons, leucine does not show a modulation effect on the Lrp-mediated regulation of *leuABCD* expression (Lin *et al.* 1992).

The monomer form of Lrp contains two protein domains: an N-terminal helix-turn-helix motif and a C-terminal antiparallel β -sheet flanked by two α -helices. The N-terminal domain allows binding of Lrp to DNA while the C-terminal domain contains a leucine-binding site. Binding of leucine to Lrp triggers a conformational alteration of Lrp, resulting in an expression regulation of the target genes through modulation of Lrp capacity to bind DNA target sites (de los Rios and Perona, 2007).

Expression of the *lrp* gene can be regulated by Lrp itself by autogenous repression (Wang *et al.*, 1994). Expression of *lrp* was also demonstrated to be dependent on the nutrients present in the cell medium: *lrp* is down-regulated in rich media, in glucose minimal media supplemented with amino acids as well as in minimal media containing alternative carbon sources, such as rhamnose, arabinose, glycerol, pyruvate, acetate or succinate (Chen *et al.*, 1997).

2.2.5.2.3.2 CAP-mediated global regulation

In addition to regulation by transcription attenuation, the *ilvBN* operon is subjected to global catabolite repression control mediated by CAP (catabolite activator protein). Freundlich (1977) suggested that the ppGpp requirement for *ilvBN* expression could be substituted by cAMP. Sutton and Freundlich (1980) demonstrated then that cAMP was indeed involved in the regulation of *ilvBN* operon, but not *ilvIH* and *ilvGM*, since cAMP allows de-repression of AHAS during valine and leucine but not isoleucine starvation. In presence of glucose (i.e. when cAMP levels are low) activity of AHAS I remains low. However, when adding cAMP in the presence of glucose or when using carbon sources different from glucose (i.e. glycerol, succinate or lactate), a 2 to 3-fold increase of AHAS I activity is reported (Sutton and Freundlich, 1980). Friden *et al.* (1982) determined the sequence of the promoter region of the *ilvBN* operon and shown that regions -35 and -72 have similarities to regulatory regions present in CRP (cAMP receptor protein)-dependent promoters. Friden *et al.* (1984, I) reported that CRP binds to the *ilvBN* promoter at a position located between -44 to -82 and prevents RNA polymerase interaction with a second non-productive binding site. Dailey and Cronan (1986) proposed that CRP-mediated activation of operon *ilvBN* might allow sufficient isoleucine and valine biosynthesis when *E. coli* grows in the presence of a carbon source different from glucose (i.e. acetate or oleate).

2.2.5.2.3.3 IHF-mediated global regulation

The global IHF (integration host factor) also regulates expression of operons *ilvBN* and *ilvGMEDA*. IHF is a DNA-binding heterodimer encoded by genes *ihfA* and *ihfB* and it plays an important role in the integration of bacteriophage lambda DNA into the *E. coli* genome, plasmid maintenance, conjugation and gene expression regulation, among others (Miller *et al.*, 1979; Friedman, 1988). Friden *et al.* (1984, II) suggested that IHF might be a positive effector for transcription of operons *ilvBN* and *ilvGMEDA* since IHF *E. coli* mutants reported decreased amounts of enzymes encoded by those operons, resulting in severe growth inhibition in minimal medium. IHF-binding sites are located at -85 in the *ilvBN* operon and at -95 and +131 in the *ilvGMEDA* operon (reviewed by Freundlich *et al.*, 1992). IHF interaction triggers expression activation of both *ilvBN* and *ilvGMEDA* operons. Tsui and Freundlich (1990) demonstrated that IHF is a direct activator of *ilvBN* transcription. When bound to *ilvBN* promoter-leader DNA region, IHF triggers DNA bending, thus acting as an antiterminator. This impedes formation of transcription termination and enables read-through of the RNA polymerase into the structural genes of operon *ilvBN*. The promoter of operon *ilvGMEDA* comprises two upstream activating sequences: UAS1 and UAS2. UAS1 contains the IHF-binding site. Binding of IHF to this site when DNA template is negatively supercoiled activates transcription of the *ilvGMEDA* operon a maximum of 5-fold (Parekh *et al.*, 1996).

2.2.5.2.3.4 Global RelA/SpoT modulon

Additionally, the global RelA/SpoT modulon is also described to regulate transcription of operons implicated in amino acid biosynthesis. Under conditions of amino acid or primary carbon source starvation, *E. coli* cells trigger a stringent response. This response is characterized by a sudden accumulation of GTP derivatives guanosine 5'-diphosphate, 3'-diphosphate (ppGpp) and guanosine 5'-triphosphate, 3'-diphosphate (pppGpp), also known as alarmones. The synthesis of (p)ppGpp is enhanced during amino acid starvation, when uncharged tRNAs bind to the acceptor site of a ribosome. The resulting stalled ribosomes activate (p)ppGpp synthetase I (RelA), which is encoded by gene *relA*. RelA remains bound to the 50S large ribosomal subunit through peptide L11 (Loveland *et al.*, 2016) and catalyzes transference of a pyrophosphoryl group from ATP to GTP to yield (p)ppGpp. An alternative RelA-independent pathway for (p)ppGpp biosynthesis is activated during long-lasting carbon starvation and requires SpoT. SpoT, encoded by gene *spoT*, is a bifunctional (p)ppGpp synthetase II/hydrolase, regulating (p)ppGpp synthesis and degradation, respectively. Alarmone levels are mainly controlled by the protein pair RelA/SpoT (termed RSH) but other enzymes such as pppGpp-5'-phosphohydrolase (Gpp, encoded by gene *gppA*) and nucleoside diphosphate kinase (Ndk, encoded by gene *ndk*) are also involved (Hauryliuk *et al.*, 2015; Lengeler *et al.*, 1999) (Figure 4).

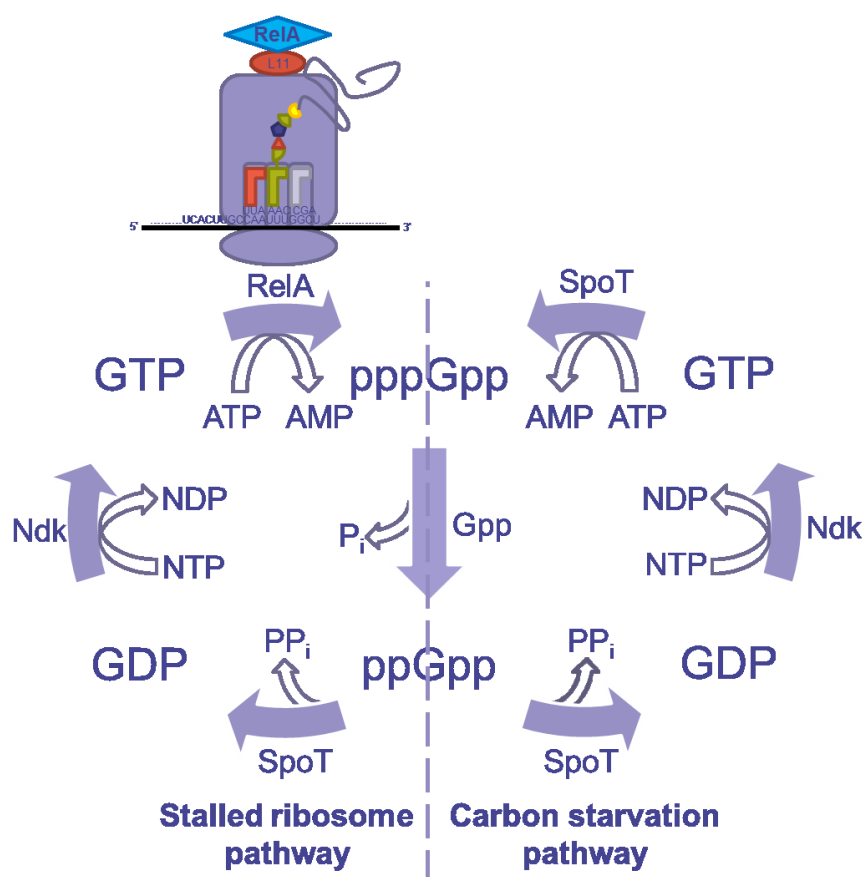


Figure 4. Main synthesis and degradation pathways of alarmone (p)ppGpp in *E. coli*. See text for description. Adapted from Lengeler *et al.* (1999).

In addition, Fang and Bauer (2018) recently found that the ACT domain present in RSH proteins from numerous bacterial species binds BCAA. This interaction results in a modulation of the alarmone synthetase/hydrolase activities, thus adding another layer of control of the stringent response. Fang and Bauer (2018) demonstrated that binding of amino acids valine and isoleucine to the ACT domain of RSH increases (p)ppGpp hydrolase activity. Accordingly, when BCAA are depleted (p)ppGpp hydrolysis would remain inhibited, thus triggering accumulation of (p)ppGpp and the subsequent stringent response.

Main effects of the (p)ppGpp-mediated stringent response comprise inhibition of RNA polymerase and rRNA/tRNA synthesis, translation repression as well as activation of biosynthetic operons, including the BCAA biosynthetic enzymes (Lengeler *et al.*, 1999; Traxler *et al.*, 2008). Artsimovitch *et al.* (2004) reported that ppGpp controls gene expression by binding to a single site of the RNA polymerase (RNAP) in two alternative orientations. Moreover, they suggested that this binding may alter structure of the RNAP active site, hence affecting NTP (nucleoside triphosphate) interactions with RNAP, and that ppGpp might also interact with the coding DNA strand in the transcription bubble.

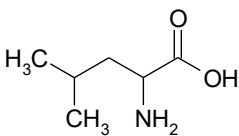
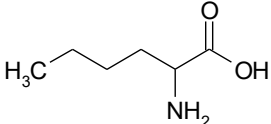
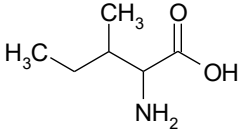
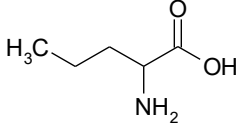
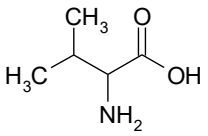
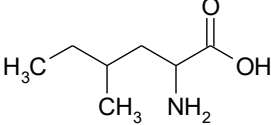
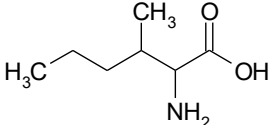
Traxler *et al.* (2008) demonstrated that genes *thrA*, *thrB* and *ilvC* as well as operons *leuABCD* and *ilvGMEDA* were significantly up-regulated under conditions of isoleucine depletion by the (p)ppGpp-mediated stringent response. Interestingly, as opposed to *ilvGM*, operons *ilvIH* and *ilvBN* did not report any significant expression modulation in that study. Other reports show that *ilvBN* operon is stimulated by ppGpp (Umbarger, 1996; Tedin and Norel, 2001) while *ilvIH* expression is actually repressed by ppGpp under conditions of amino acid starvation (Baccigalupi *et al.*, 1995).

2.3 Non-canonical branched chain amino acids (ncBCAA)

2.3.1 Introduction to ncBCAA

A branched-chain amino acid (BCAA) is an amino acid containing an aliphatic side-chain. The proteinogenic or canonical BCAA are leucine, isoleucine and valine. The most prominent non-proteinogenic or non-canonical BCAA include norleucine, norvaline, homoisoleucine and β -methylnorleucine. The structure of the different BCAA is shown in Table 4.

Table 4. Molecular structure of canonical and non-canonical BCAA

Canonical BCAA		Non-canonical BCAA	
Amino acid	Molecular structure	Amino acid	Molecular structure
Leucine		Norleucine	
Isoleucine		Norvaline	
Valine		Homoisoleucine	
		β -methylnorleucine	

The non-canonical BCAA can be produced by the *E. coli* metabolism as byproducts of the BCAA biosynthetic pathway under particular conditions. These modified amino acids can be secreted to the medium and/or mis-incorporated into cellular proteins through tRNA misaminoacylation during protein translation. Such mis-incorporation can lead to the production of altered proteins, having non optimal characteristics e.g. altered biological activity, modulated sensitivity to proteolysis and immunogenicity (Laird and Veeravalli, 2013). Hence, it has become a crucial matter of contention in the pharmaceutical industry since product quality (high purity and homogeneity) is pivotal for commercial approval of recombinant proteins that are to be used as human therapeutics (Apostol *et al.*, 1997). However, non-canonical amino acids, including ncBCAA, also awakens interest in the field of synthetic biology since directed mis-incorporation of these translationally active analogues opens the possibility to develop proteins with new functionalities. The main strategy employed to trigger mis-incorporation of a given non-canonical amino acid in positions corresponding to the canonical

counterpart (residue-specific incorporation of non-canonical amino acids) is the supplementation of that non-canonical amino acid into the cultivation medium while using an *E. coli* strain auxotrophic for the canonical counterpart. Under those conditions, *E. coli* is forced to utilize the supplemented non-canonical amino acid for protein translation (Kiick *et al.*, 2001). This strategy allows then the global replacement of a given residue by a non-canonical counterpart. However, there are also strategies enabling replacement of only a single residue in a certain position of the protein (site-specific incorporation of non-canonical amino acids). The traditional strategy consists of the supplementation of chemically aminoacylated suppressor tRNA with the non-canonical amino acid to an *in vitro* cell-free system where target codon of the gene encoding the recombinant protein is substituted by an amber nonsense codon (TAG) (reviewed by Link *et al.*, 2003). Recently, a new strategy was developed enabling incorporation of (synthetic) non-canonical amino acids into recombinant proteins in living cell systems. This includes the transference of an orthogonal translation system including the aminoacyl-tRNA synthetase (aaRS) and the tRNA pair corresponding to the non-canonical amino acid of interest as well as substitution of the target codon by a dedicated one (reviewed by Young and Schultz, 2010).

Among the non-canonical BCAA, norleucine is the one which accumulates more literature knowledge. The first evidence of the incorporation of exogenous norleucine into a recombinant protein by *E. coli* dates from 1956 (Munier and Cohen, 1956). Then, different studies reported that the mis-incorporation of exogenous norleucine into recombinant proteins by *E. coli* took place at positions where methionine is normally incorporated (Cohen and Munier, 1959; Cowie *et al.*, 1959), thus confirming that norleucine is a structural analog of methionine. From that period onwards plenty of literature was made available demonstrating exogenous norleucine mis-incorporation into a wide range of recombinant proteins by the *E. coli* production platform in methionine positions, including recombinant adenylate kinase (Gilles *et al.*, 1988), recombinant mammalian calmodulin (Yuan and Vogel, 1999) and recombinant cytochrome P450 BM-3 heme domain (Cirino *et al.*, 2003). There were also cases reported where norleucine was not being supplied exogenously in the media, but was being naturally synthesized in *E. coli* cells and incorporated into recombinant proteins. In that case norleucine was found to be incorporated into recombinant interleukin-2 (IL-2) (Lu *et al.*, 1988), recombinant bovine somatotropine (bST) (Bogosian *et al.*, 1989), recombinant human macrophage colony stimulating factor (hM-CSF) (Randhawa *et al.*, 1994), recombinant human brain-derived neurotrophic factor (Sunasara *et al.*, 1999) and in a 41 kDa Met-rich recombinant protein vaccine candidate (Ni *et al.*, 2015). Despite all cases described for recombinant proteins, there are no evidences in the literature regarding the synthesis and incorporation of norleucine into natural non-recombinant proteins by *E. coli*. However, the natural presence of norleucine has been reported in the field-growing parasitic fungi *Claviceps purpurea* (Cvak *et al.*, 2005).

Formation of norleucine by complex regulatory mutants of *Serratia marcescens* has been reported (Kisumi *et al.* 1976, I and II). The expression of recombinant leucine-rich proteins may trigger the biosynthesis of norleucine in *E. coli* (Bogosian *et al.*, 1989). Norleucine can be incorporated into proteins both at internal residues as well as the amino terminus and the incorporation at the methionine loci in the protein was demonstrated to be random (Bogosian *et al.*, 1989). However, recent evidences reveal that, at certain positions, higher levels of norleucine incorporation are reported (Veeravalli *et al.*, 2015).

Norvaline was initially reported as a natural component of an antifungal peptide manufactured in *Bacillus subtilis* (Nandi and Sen, 1953). Norvaline was identified as a byproduct in isoleucine

overproducing regulatory mutants of *Serratia marcescens* (Kisumi *et al.*, 1976, I and II). Similarly to norleucine, the incorporation of norvaline into recombinant proteins at leucine positions can be intentionally triggered by feeding microorganisms with exogenous norvaline (Miyazawa *et al.*, 1989). In addition, norvaline can also be naturally synthesized in *E. coli* cells and incorporated into recombinant proteins (recombinant hemoglobin, Apostol *et al.*, 1997). Recent experimental results have shown that the *in vivo* production of norvaline by *E. coli* is boosted in conditions of oxygen limitation (Soini *et al.*, 2008, I).

Additionally, other reports indicate that the non-canonical amino acids β -methylnorleucine (recombinant hirudin, Muramatsu *et al.*, 2002) and homoisoleucine (recombinant human brain-derived neurotrophic factor, Sunasara *et al.*, 1999) are sometimes inappropriately inserted into heterologous proteins, in the place of isoleucine. β -methylnorleucine was first discovered to be produced by *Serratia marcescens* (Sugiura *et al.*, 1981). Biosynthesis of homoisoleucine in *Serratia marcescens* is described in Kisumi *et al.* (1976, I and II). The incorporation of homoisoleucine into the recombinant coiled-coil peptide A1 was induced by homoisoleucine addition to the medium by Van Deventer *et al.* (2011).

2.3.2 ncBCAA biosynthetic pathway

The initial outline for norleucine, norvaline and homoisoleucine biosynthesis was proposed by Kisumi *et al.* (1976, I and II) by using regulatory mutants of leucine biosynthesis in *Serratia marcescens*. The scheme for β -methylnorleucine biosynthesis was later suggested by Sugiura *et al.* (1981). Some evidences also indicate that *E. coli* utilize the same metabolic pathway than *Serratia marcescens* to synthesize non-canonical branched-chain amino acids (Tsai *et al.*, 1988; Bogosian *et al.*, 1989; Muramatsu *et al.*, 2003), confirming that it derives from the isoleucine route.

A schematic representation of the ncBCAA biosynthetic pathway is shown in Figure 5. The first substrate of both isoleucine and ncBCAA biosynthetic pathways is α -ketobutyrate, which can be mainly synthesized from the deamination of threonine by threonine deaminase (*ilvA*) but also from pyruvate via keto acid chain elongation by the leucine operon (*leuABCD*) (Bogosian *et al.*, 1989). α -ketobutyrate can be either transformed to isoleucine by *ilv* operons or sequentially converted to α -ketovalerate via keto acid chain elongation by the leucine operon (*leuABCD*). In detail, α -isopropylmalate synthase condensates the acetyl group of acetyl-CoA with α -ketobutyrate to produce α -ethylmalate and CoA. Then, α -isopropylmalate isomerase catalyzes isomerization of α -ethylmalate to β -ethylmalate. Later, the NAD-dependent β -isopropylmalate dehydrogenase catalyzes the oxidative decarboxylation of β -ethylmalate to produce α -ketovalerate. α -ketovalerate can, in turn, be either transformed to α -keto- β -methylcaproate by *ilv* operons, either sequentially converted to α -ketocaproate via keto acid chain elongation by the leucine operon (*leuABCD*) or subjected to transamination by transaminase B to yield norvaline (Tsai *et al.*, 1988; Bogosian *et al.*, 1989; Sycheva *et al.*, 2007; Soini *et al.*, 2008, I).

Keto acid chain elongation of α -ketovalerate by the leucine operon (*leuABCD*) is described in detail below. α -isopropylmalate synthase condensates the acetyl group of acetyl-CoA with α -ketovalerate to produce α -propylmalate and CoA. Then, α -isopropylmalate isomerase catalyzes isomerization of α -propylmalate to β -propylmalate. Later, the NAD-dependent β -isopropylmalate dehydrogenase catalyzes the oxidative decarboxylation of β -propylmalate to produce α -ketocaproate. Transaminase

B catalyzes then transamination of α -ketocaproate to yield norleucine (Tsai *et al.*, 1988; Bogosian *et al.*, 1989; Sycheva *et al.*, 2007; Soini *et al.*, 2008, I).

Transformation of α -ketovalerate by *ilv* operons is shown in detail next. First, acetohydroxy acid synthase catalyzes a 2-step bi-substrate enzymatic reaction. In the first reaction step pyruvate binds to AHAS, undergoing decarboxylation to form an active intermediate. In the second reaction step α -ketovalerate binds to the pre-formed intermediate to generate α -aceto- α -hydroxyvalerate. Second, acetohydroxy acid isomeroreductase catalyzes the conversion of α -aceto- α -hydroxyvalerate to α,β -dihydroxy- β -methylcaproate. Third, dihydroxyacid dehydratase drives dehydration of the α,β -dihydroxy acid intermediate to yield α -keto- β -methylcaproate. Fourth, transaminase B catalyzes transamination of α -keto- β -methylcaproate to form β -methylnorleucine by using glutamate as donor of the amino group (Muramatsu *et al.*, 2003).

Homoisoleucine can be also synthesized by the keto acid chain elongation reaction catalyzed by the *leu* operon (*leuABCD*) and the subsequent transamination of α -keto- β -methylvalerate, a metabolic intermediate of the isoleucine biosynthetic route (Kisumi *et al.*, 1976, I and II).

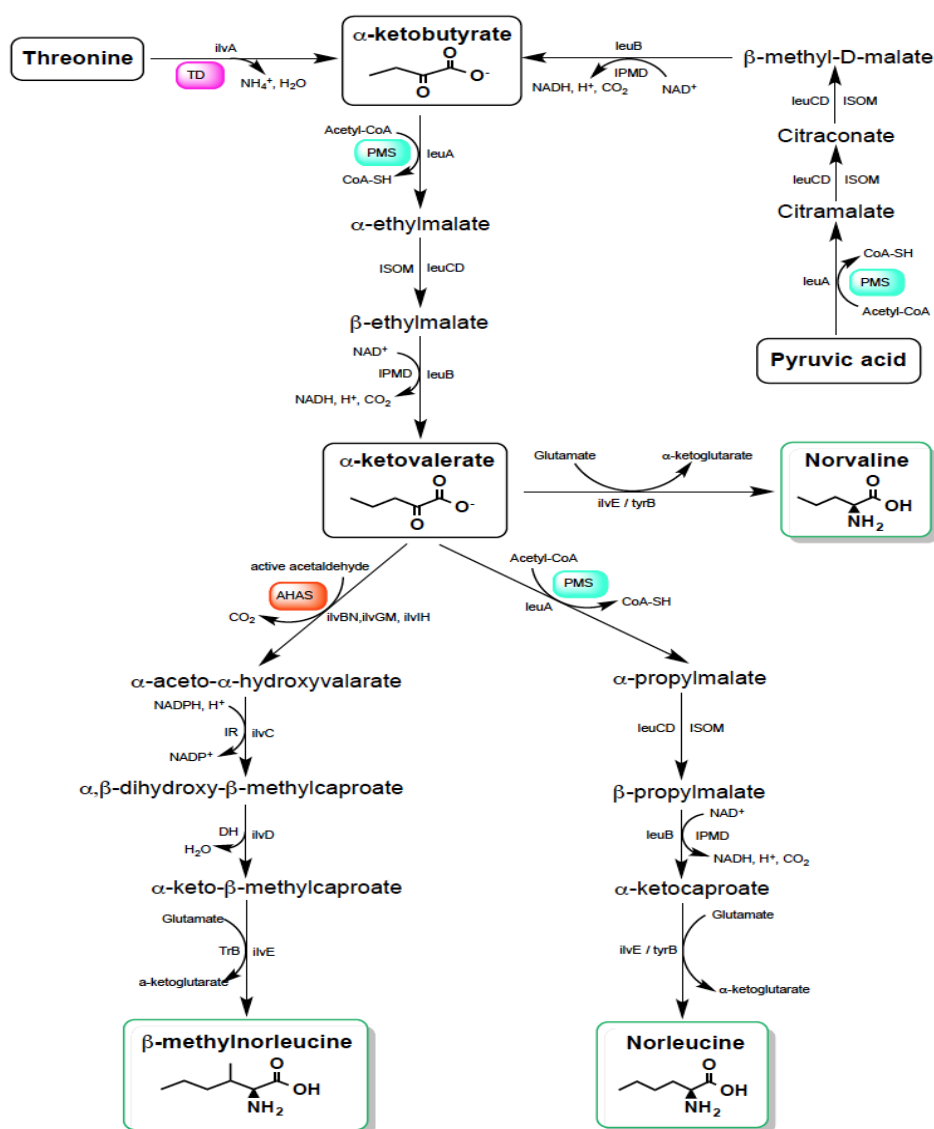


Figure 5. Schematic representation of the ncBCAA biosynthetic pathway. Gene and enzyme names displayed in this figure are the same than the ones described in Figure 4. See text for description. Adapted from Reitz (2017).

2.3.3 Substrate promiscuity of enzymes involved in ncBCAA biosynthesis

The synthesis and accumulation of non-canonical BCAA results from the low specificity of the *leu* and *ilv*-operon-coded enzymes involved in the BCAA biosynthetic pathway for their substrates. This explains the sequential keto acid chain elongation from pyruvate to α -ketocaproate over α -ketobutyrate and α -ketovalerate by the actuation of the *leu* enzymes α -isopropylmalate synthase (*leuA*), β -isopropylmalate dehydrogenase (*leuB*) and α -isopropylmalate isomerase (*leuCD*). Despite being α -ketoisovalerate the preferred substrate for α -isopropylmalate synthase, this enzyme also shows certain affinity towards alternative α -keto acids such as pyruvate, α -ketobutyrate and α -ketovalerate. Table 5 shows kinetic parameters of α -isopropylmalate synthase with various substrates. Since kinetic information of *E. coli*- α -IPMS was not found in the literature, kinetic parameters of α -IPMS from other organisms were considered.

Table 5. Kinetic parameters of α -IPMS from different microorganisms towards various α -ketoacids. The identity percentage of the α -IPMS of a certain microorganism with respect to *E. coli*- α -IPMS is shown.

Organism	Identity <i>E. coli</i>	Parameter	Substrate				Ref.
			α -KIV	Pyruvate	α -KB	α -KV	
<i>Salmonella thyphimurium</i>	93 %	K_m (mM)	0.06	10	1.1	-	Kohlhaw <i>et al.</i> (1969)
<i>Alcaligenes eutrophus</i>	49 %	K_m (mM)	0.06	10	1.8	0.4	Wiegel and Schlegel (1977)
<i>Mycobacterium tuberculosis</i>	29 %	K_m (mM)	0.03	27.3	1.1	0.8	Hunter and Parker (2014)
		k_{cat} (s^{-1})	0.81	0.49	0.84	0.29	
		k_{cat}/K_m ($mM^{-1}s^{-1}$)	27	0.018	0.764	0.363	
<i>Serratia marcescens</i>	24 %	K_m (mM)	0.77	3.4	7.7	9	Kisumi <i>et al.</i> (1976, I)

Moreover, the three consecutive reactions catalyzed by the *ilv* enzymes acetohydroxy acid synthase (*ilvBN*, *ilvGM*, *ilvIH*), acetohydroxy acid isomeroreductase (*ilvC*) and dihydroxyacid dehydratase (*ilvD*) can take place in parallel in the biosynthetic pathways of valine, isoleucine and β -methylnorleucine. As aforementioned, acetohydroxy acid synthase catalyzes a 2-step enzymatic reaction. In the first reaction step only pyruvate can bind to AHAS, catalyzing decarboxylation to form an active intermediate. In the second reaction step three alternative substrates (pyruvate, α -ketobutyrate or α -ketovalerate) can bind to the pre-formed intermediate to generate three alternative acetohydroxy acids (Gollop *et al.*, 1989). Pyruvate is the preferred substrate for acetohydroxy acid synthase, but it is also able to transform the other two. Nevertheless, the complexity of acetohydroxy acid synthase specificity gets even higher considering that there are three different isoenzymes available, each one showing different substrate preference and catalytic properties (Table 6).

Table 6. Kinetic and affinity parameters of the three acetohydroxyacid synthase isoenzymes (AHAS I, II and III) towards various substrates. *The factor R1 characterizes the specificity of enzyme acetohydroxyacid synthase for α -acetohydroxybutyrate formation in relation to α -acetolactate formation. R1 is defined as $R1 = (v_{AHB}/v_{AL})/([2\text{-ketobutyrate}]/[\text{pyruvate}])$. The factor R2 characterizes the specificity of enzyme acetohydroxyacid synthase for α -aceto- α -hydroxyvalerate formation in relation to α -acetolactate formation. R2 is defined as $R2 = (v_{AHV}/v_{AL})/([2\text{-ketovalerate}]/[\text{pyruvate}])$. a: AHAS from *E. coli* K-12; b: AHAS from *S. typhimurium*.

Parameter	AHAS I	AHAS II	AHAS III	Reference
K_m (pyruvate) (mM)	1.5 ^a	10.5 ^b	6 ^a	Reviewed by Gollop <i>et al.</i> (1989)
k_{cat}/K_m (pyruvate) (mM ⁻¹ s ⁻¹)	14 ^a	6.6 ^a	0.36 ^a	Reviewed by Vinogradov <i>et al.</i> (2006)
R1*	2 ^a	65 ^b	40 ^a	Reviewed by Gollop <i>et al.</i> (1989)
R2*	<0.1 ^a	2.4 ^b	2.3 ^a	Reviewed by Gollop <i>et al.</i> (1989)

Transaminase B (*ilvE*) also reports substrate promiscuity since it catalyzes transamination of leucine, valine, isoleucine, norleucine, norvaline, β -methylnorvaline and homoisoleucine intermediates (Table 7).

Table 7. Kinetic parameters of transaminase B from *E. coli* towards various substrates. Data obtained from Yu *et al.* (2014).

Substrate	Product	K_m (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_m (mM ⁻¹ s ⁻¹)
α -ketoisocaproate	leucine	0.08	24.7	309
α -ketoisovalerate	valine	0.20	10.5	52.5
α,β -keto- β -methylvalerate	isoleucine	0.07	23.0	329
α -ketovalerate	norvaline	0.60	25.9	43.2
α -ketocaproate	norleucine	0.22	23.1	105

2.3.4 Conditions triggering ncBCAA biosynthesis

According to what has been described in the literature so far, three are the main causes of ncBCAA production in *E. coli*-based recombinant protein production processes: overflow metabolism and mixed-acid fermentation, enzymatic feedback regulation and activity of acetohydroxyacid synthase isoenzyme II.

2.3.4.1 Overflow metabolism and mixed-acid fermentation

The scale-up of recombinant protein production processes from laboratory scale to large-scale reactors often lead to numerous problems. For instance, for an *E. coli*-based recombinant protein production process a 20% reduction of biomass yield and an increase of by-product formation was reported by Bylund *et al.* (1998) when scaling up from 3L to 9m³. Riesenber *et al.* (1990) not only showed a biomass reduction but also a lower product yield when scaling up an *E. coli* process from 30L to 450L. During fermentation, gradient zones of substrate, dissolved oxygen, pH and other parameters are formed due to inefficient mixing and *E. coli* cells respond to these environmental changes by modulating their metabolism (Schweder *et al.*, 1999). The standard procedure to grow *E. coli* for recombinant protein production processes at large-scale consists of a fed-batch operation under aerobic conditions. Addition of concentrated feeding solution is normally carried out from the top of the reactor while aeration is done by providing oxygen through a vent located at the bottom of the reactor. Due to the increased mixing times occurring at large-scale, provided glucose and oxygen are distributed in the reactor in an inverse gradient manner. Hence, reactor regions close to the glucose inlet are characterized by high glucose concentrations and low dissolved oxygen while reactor areas close to the oxygen vent show glucose depletion and high dissolved oxygen. Central regions of the reactor report intermediate values of glucose and oxygen if compared with upper and lower areas. Moreover, oxygen limitation in the upper part is exacerbated because cells increase oxygen consumption under glucose excess. *E. coli* responds to glucose excess and oxygen limitation by shifting metabolism from oxidative respiration to mixed-acid fermentation, resulting in overflow metabolism (Enfors *et al.*, 2001). Under these conditions, not only the mixed-acid fermentation products accumulate, but also pyruvate (Soini *et al.*, 2008, I). Pyruvate excess present intracellularly increases the metabolic flux going to ncBCAA biosynthesis through the sequential keto acid chain elongation from pyruvate to α -ketocaproate over α -ketobutyrate and α -ketovalerate by the actuation of the *leu* operon-encoded enzymes (Apostol *et al.*, 1997). This hypothesis is supported by the observations reported by Soini *et al.* (2011): the combination of oxygen limitation with a constant glucose supply in a two-compartment STR-PFR scale-down reactor reported a significant impact on enhancing norvaline biosynthesis due to pyruvate accumulation in a recombinant *E. coli* cultivation. Furthermore, Soini *et al.* (2008, I) originally reported accumulation of pyruvate-based amino acids such the ncBCAAs norleucine and norvaline as well as alanine and valine in a standard STR fed-batch *E. coli* cultivation under glucose excess and induced oxygen limitation upon a stirrer downshift. It was proposed that a strong accumulation of pyruvate is a prerequisite for α -ketobutyrate formation since α -IPMS shows lower affinity towards pyruvate than the other alternative substrates: α -ketoisovalerate and α -ketobutyrate (Table 5). Similarly, α -ketobutyrate has been proposed as a prerequisite for ncBCAA biosynthesis, since the affinity of α -IPMS for α -ketobutyrate is around 20-fold lower compared to its preferred substrate, α -ketoisovalerate (Table 5) (Sycheva *et al.*, 2007).

2.3.4.2 Feedback regulation

Another factor strongly influencing ncBCAA biosynthesis is the de-regulation of the *leuABCD* operon. It encodes the enzymes catalyzing the chain elongation of various α -keto acids, generating precursors for leucine and ncBCAAs biosynthesis. Chain elongation of α -ketoisovalerate to α -ketoisocaproate directs to leucine biosynthesis. Consecutive chain elongation of pyruvate to α -ketocaproate, via α -

ketobutyrate and α -ketovalerate ends up with ncBCAA biosynthesis. α -ketovalerate and α -ketocaproate are the precursors of norvaline and norleucine, respectively. Sycheva *et al.* (2007) demonstrated that an up-regulation of the *leuABCD* operon triggers an increase of norvaline and norleucine accumulation. Furthermore, transcription of the *leu* operon is negatively regulated by the presence of leucine, either via attenuation or via Lrp-regulation (Vitreschak *et al.*, 2006; Lin *et al.*, 1992). Enzymes encoded by the *leu* operon are also subjected to leucine-mediated feedback inhibition (Soper *et al.*, 1976). Accordingly, when leucine is in excess the *leuABCD* operon is repressed, thus limiting leucine and ncBCAA biosynthesis. However, conditions leading to lower intracellular levels of leucine cause a de-repression of the *leu* operon, hence boosting the metabolic flux going to leucine and ncBCAA biosynthetic pathways. These conditions are especially accomplished in *E. coli*-based leucine-rich recombinant protein production processes, where the metabolic demand for leucine is enhanced (Bogosian *et al.*, 1989; Fenton *et al.*, 1994; Apostol *et al.*, 1997). This is of special interest since numerous recombinant proteins present a high leucine-content (Reitz *et al.*, 2018). For instance, recombinant insulin used in this thesis has a leucine content of 14.5% while an average *E. coli* protein only contains 8.4%. In addition to the de-regulation of the *leu* operon, ncBCAA biosynthesis requires accumulation of α -ketobutyrate as the affinity of α -IPMS for α -ketobutyrate is lower compared to its preferred substrate, α -ketoisovalerate (Table 5) (Sycheva *et al.*, 2007).

2.3.4.3 Activity of acetohydroxyacid synthetase

Biermann *et al.* (2013) reported that the degree of norleucine and norvaline biosynthesis strongly relies on the *E. coli* strain selected as recombinant protein expression system since strain *E. coli* BL21(DE3) reported a significant low accumulation of ncBCAA with respect to strain *E. coli* K-12. As opposed to strain *E. coli* BL21(DE3), genome sequencing and annotation studies revealed that a two-base insertion event between base pairs 1250 and 1253 is present in the coding sequence of gene *ilvG* in *E. coli* K-12 strains. This insertion causes a shift of the reading frame and, as a consequence, a stop codon is formed, resulting in a premature termination of *ilvG* gene expression (Yoon *et al.*, 2012; Parekh and Hatfield, 1997; Lawther *et al.* 1981). AHAS II is encoded by *ilvG* and it is involved in biosynthesis of isoleucine starting from α -ketobutyrate and, in contrast to AHAS I and AHAS III, AHAS II is resistant to valine-mediated feedback inhibition. The absence of AHAS II activity in *E. coli* K-12 strains leads to valine toxicity: AHAS I and AHAS III are inhibited in the presence of valine and, since AHAS II cannot be properly synthesized, a drastic reduction of leucine and isoleucine biosynthesis occurs. Under these conditions, growth behavior is jeopardized (Anderson *et al.* 2001, Biryukova *et al.* 2010) and accumulation of α -ketobutyrate is enhanced, thus resulting in higher ncBCAA production (Soini *et al.*, 2008, I; Sycheva *et al.*, 2007).

2.3.5 Mechanism of ncBCAA mis-incorporation into proteins during translation

Aminoacyl-tRNA synthetases (aaRSs) catalyze the transference of amino acids to their corresponding cognate tRNA in 2 steps. The first step consists of an ATP-mediated amino acid activation yielding an aminoacyl-adenylate intermediate. In the second step, the aminoacyl moiety from the activated intermediate is transferred to the cognate tRNA thus yielding an aminoacylated tRNA, which can

then be transferred to the ribosome by the elongation factor Tu (EF-Tu) for subsequent protein translation (Cvetesic *et al.*, 2012).

The fidelity of protein synthesis counts on the aptitude of aaRSs to charge the appropriate canonical amino acid onto its corresponding tRNA. Such fidelity can be jeopardized by a number of non-canonical amino acids, particularly ncBCAA, which are structurally similar to their canonical equivalents (Martinis *et al.*, 1997). Promiscuity of aaRSs can be appreciated in Table 8, where kinetic parameters of leuRS toward different amino acids are shown (information extracted from Tang and Tirrell, 2002).

Table 8. Kinetic parameters of leuRS from *E. coli* towards various amino acids. Data obtained from Tang and Tirrell (2002).

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)
leucine	0.018	2.2	122.22
isoleucine	2.568	0.06	0.023
valine	2.356	0.03	0.013
methionine	2.178	0.08	0.037
norvaline	1.155	1.24	1.074
norleucine	2.516	0.22	0.087

For instance, leucyl-tRNA synthetase (leuRS) must distinguish between leucine and the non-canonical counterpart norvaline, which only differ by a single methyl group (Apostol *et al.*, 1997) (Table 4). Homoisoleucine can also compete with leucine for leuRS (Van Deventer *et al.*, 2011). The same happens with methionyl-tRNA synthetase (metRS), which must discriminate between methionine and norleucine (Kiick *et al.*, 2001), and isoleucyl-tRNA synthetase (ileRS), which must differentiate between isoleucine and β -methylnorleucine (Muramatsu *et al.*, 2003). In order to prevent mis-incorporation of non-canonical amino acids, aaRSs have evolved quality control mechanisms which allow hydrolysis of misactivated amino acids (pre-transfer editing) or misacylated tRNAs (post-transfer editing) (Figure 6). For leuRS, minincorporation is mainly prevented by post-transfer editing while for ileRS pre-transfer editing mechanism is preferred (Chen *et al.*, 2011; Cvetesic *et al.*, 2014). However, the efficiency of the editing mechanism is controlled by kinetic partitioning between synthetic and editing reaction. For instance, when kinetic partitioning of norvalyl-tRNA^{Leu} between post-transfer editing hydrolysis at the LeuRS editing site and dissociation from leuRS is altered, an accumulation of norvalyl-tRNA^{Leu} might occur, hence increasing mis-incorporation of norvaline at leucine positions by the translation machinery, since elongation factor Tu (EF-Tu) does not discriminate against norvaline (Cvetesic *et al.*, 2012). In addition, Apostol *et al.* (1997) shown that the ratio of free norvaline with respect to free leucine strongly determines the degree of norvaline mis-incorporation in the recombinant protein sequence at leucine positions. Hence, under cultivation conditions triggering biosynthesis of ncBCAA, those accumulate in the cell so that ratio of ncBCAAs relative to their canonical counterparts increases and this might alter kinetic partitioning at the editing site of aaRSs, resulting in a reduced editing efficiency and, as a consequence, a major ncBCAA mis-incorporation. Tang and Tirrell (2002) revealed importance of editing mechanisms in preventing

ncBCAA mis-incorporation since when editing capacity is attenuated, mis-incorporation of norvaline is dramatically boosted.

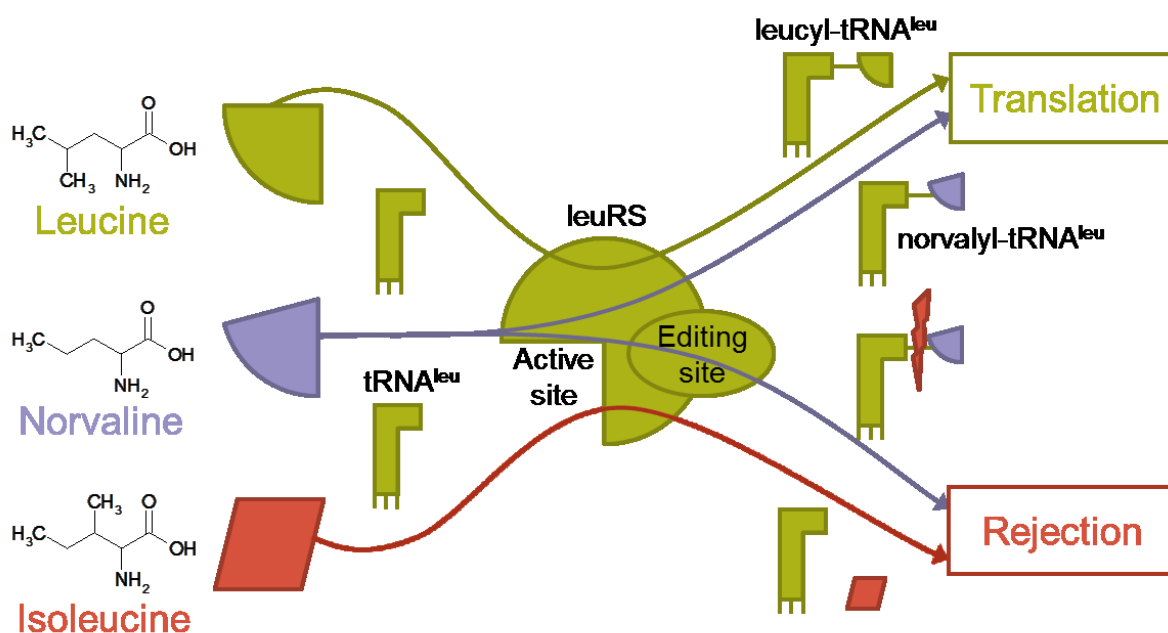


Figure 6. Strategies of leuRS to discriminate among leucine, isoleucine and norvaline. Isoleucine cannot be processed by leuRS due to its low affinity towards it. However, other amino acids such as norvaline can be identified by leuRS instead of leucine. Mischarging of amino acids onto tRNA can be prevented by proofreading at the editing site located close to the active site of aaRS. In most cases mischarged tRNA can be hydrolysed at editing site of aaRS but, under particular conditions, editing mechanisms work inefficiently, hence triggering mis-incorporation during translation. Adapted from Cvetesic *et al.* (2013) and de Pouplana (2014).

2.3.6 Inconvenients of ncBCAA biosynthesis and mis-incorporation into recombinant proteins

Biosynthesis of ncBCAA can have negative effects in the physiology of *E. coli*. For instance, norleucine inhibits DNA replication and methylation reactions, causing DNA damage. In addition, norleucine restricts methionine biosynthesis by inhibiting homoserine succinyltransferase, the first enzyme of the methionine biosynthetic pathway (summarized by Bogosian *et al.*, 1989).

Mis-incorporation of ncBCAA during translation can lead to the production of altered proteins, having non optimal characteristics e.g. altered biological activity, modulated sensitivity to proteolysis and immunogenicity (Laird and Veeravalli, 2013). There are plenty of evidences showing that the mis-incorporation of norleucine into recombinant proteins by *E. coli* may lead to alteration of protein structure and biological properties. Recombinant norleucine-substituted β -galactosidase was more resistant to alkylation (Naider *et al.*, 1972). Recombinant norleucine-rich adenylate kinase shows a much higher resistance to hydrogen peroxide inactivation under denaturing conditions than the methionine-rich variant (Gilles *et al.*, 1988). Recombinant norleucine-rich mammalian calmodulin (CaM) presents a decreased enzymatic activity if compared to the normal variant (Yuan and Vogel, 1999). Recombinant norleucine-rich cytochrome P 450 BM-3 heme domain shows an increased peroxxygenase activity than its normal methionine-containing counterpart (Cirino *et al.*, 2003).

When recombinant proteins are to be used as pharmaceutical products for human use, a number of quality criteria have to be fulfilled in order to ensure its effectiveness and safety when delivered in the market. Hence, a number of product parameters have to be tested and these should meet the specifications. One of those parameters is the purity level of the product. The mis-incorporation of ncBCAA into the product during the recombinant protein production process would generate a pool of protein variants differing by only a few amino acids. During downstream processing removal of undesired variants should be performed in order to satisfy the purity specifications. However, this is sometimes very difficult since most protein variants would show similar properties and sensitivity of the analytic methods employed might not be enough. This issue represents an important concern for the pharmaceutical industry since undesired protein variants might have negative effects in patients (Harris and Kilby, 2014).

2.3.7 Strategies to avoid ncBCAA mis-incorporation into recombinant proteins

Different strategies have been used in order to reduce the degree of non-canonical BCAA mis-incorporation into recombinant proteins expressed in *E. coli*, especially for the case of norleucine: (1) mutating methionine codons of the gene that encodes the recombinant protein so that no methionine residues are present (Brunner *et al.*, 1997), (2) co-expressing enzymes that are capable of degrading non-canonical BCAA (Bogosian *et al.*, 2013), (3) supplementing the cultivation medium with exogenous canonical amino acids to reduce the likelihood of the non-canonical counterpart to be selected by the corresponding aminoacyl-tRNA synthetase (Apostol *et al.*, 1997; Tsai *et al.*, 1988; Bogosian *et al.*, 1989; Fenton *et al.*, 1994; Randhawa *et al.*, 1994; Brunner *et al.*, 1997; Abu-Absi *et al.*, 2008), (4) overproducing methionine by mutating genes involved in methionine biosynthesis and regulation such as *metA*, *metK* and *metJ* (Usuda and Kurahashi, 2005; Veeravalli *et al.*, 2015; Laird and Veeravalli, 2013) and in threonine biosynthesis such as *thrB* and *thrC* (Usuda and Kurahashi, 2005), (5) knocking-out genes that are involved in the non-canonical BCAA biosynthetic pathway such as *ilvA* and *leu* operon (Tsai *et al.*, 1988; Bogosian *et al.*, 1989; Fenton *et al.*, 1994), (6) optimizing operational conditions during fermentation (Ni *et al.*, 2015), (7) supplementing trace elements molybdenum, selenium and nickel in the cultivation medium (Biermann *et al.*, 2013) and (8) using alternative *E. coli* expression strains less prone to non-canonical BCAA mis-incorporation (Ni *et al.*, 2015).

However, most of these strategies present numerous disadvantages. Substitution of methionine codons for other ones eliminates possibility of norleucine to be mis-incorporated into the recombinant protein. Nevertheless, alteration of the native amino acid sequence of the target protein might cause structural changes in the protein conformation leading to reduction of the protein activity.

Co-expression of ncBCAA-degrading enzymes represents a powerful tool to reduce mis-incorporation of ncBCAA since no cBCAA supplementation is necessary. It was demonstrated that wild type and mutant versions of phenylalanine dehydrogenase, valine dehydrogenase, leucine dehydrogenase and glutamate dehydrogenase from numerous species of bacteria and other organisms are capable of degrading ncBCAA via NAD⁺-dependent oxidative deamination (reviewed by Bogosian *et al.*, 2013), which consists in removing the alpha-amino group from amino acids, generating then α -ketoacids, ammonia and NADH. However, these enzymes are not only specific for the ncBCAA but are also able

to degrade the canonical counterparts to a certain extent, which can lead to lower recombinant protein yields. Hence, in order to make this strategy feasible, more studies concerning protein engineering should be done in order to increase specificity of the degrading enzymes against the ncBCAA.

The strategy based on canonical amino acid supplementation works effectively to ensure product quality since it dramatically reduces the likelihood of ncBCAA to be recognised by the corresponding aminoacyl-tRNA synthetase. Moreover, addition of leucine not only reduces mis-incorporation of norvaline due to a higher competitiveness of leucine for leucyl-tRNA synthetase but also a reduction of norleucine mis-incorporation, since leucine represses the *leu* operon at both transcriptional and enzymatic level, hence reducing the metabolic flux going to ncBCAA synthesis. However, this approach is only feasible at small-scale since in industrial scale production processes additional medium supplementation would excessively increase process costs and it might also hinder operation. In addition, media supplementation would dilute fermentation content thus having a negative effect on cell density and product yield.

Overproduction of methionine by mutated versions of *metA*, *metK* and *metJ* genes overcomes all aforementioned inconvenients since it was demonstrated that growth behavior and recombinant protein production yields in some of the mutant strains are close to the control non-modified strain (Laird and Veeravalli, 2013). This strategy consists in generating new *E. coli* strains which are feedback resistant against methionine and S-adenosylmethionine (SAM), which allosterically inhibit activity of the methionine biosynthetic enzyme encoded by *metA* (homoserine transsuccinylase). A substitution of tyrosine by cysteine at position 294 (Y294C) in the amino acid sequence of homoserine transsuccinylase resulted in a feedback-resistant enzyme (Laird and Veeravalli, 2013). The enzyme encoded by *metK* (methionine adenosyltransferase) transforms methionine into SAM. Both mutation consisting of a substitution of valine by glutamic acid at amino acid position 185 (V185E) and mutation consisting of a deletion of cytosine at base position 1132 in the DNA sequence of *metK* resulted in a partially non-functional methionine adenosyltransferase (Laird and Veeravalli, 2013). These new *metK* variants would result in low levels of SAM and consequently higher de-repression of the methionine biosynthetic enzymes encoded by *metA*, *metB*, *metC* and *metH* or *metE*. MetJ (encoded by *metJ*) is a repressor regulatory protein which, when bound to its co-repressor SAM, represses methionine biosynthetic enzymes at transcriptional level. Both mutation consisting of a substitution of serine by asparagine at amino acid position 54 (S54N) (Nakamori *et al.*, 1999) of MetJ and mutation consisting of a complete *metJ* disruption (Usuda and Kurahashi, 2005) resulted in de-repression of methionine biosynthetic enzymes and methionine overproduction. Moreover, Usuda and Kurahashi (2005) reported an increased methionine excretion in *E. coli* when using a strain where threonine biosynthetic genes *thrB* and *thrC* were disrupted from the genome. Homoserine kinase (*thrB*) catalyzes phosphorylation of homoserine to homoserine phosphate while threonine synthetase (*thrC*) converts homoserine phosphate to threonine. Since homoserine transsuccinylase (*metA*) and homoserine kinase (*thrB*) compete for homoserine, hence a threonine auxotrophic strain is expected to trigger methionine overproduction. Engineered *E. coli* strains containing one or a combination of mentioned *metA*, *metK*, *metJ* and *thrBC* mutations would then result in a significant reduction of norleucine mis-incorporation. Despite the many advantages of this strategy with respect to previous ones, it is noteworthy to highlight that through methionine overproduction only norleucine mis-incorporation would be reduced while norvaline and β -methylnorleucine would still represent a problem for product quality.

L-threonine dehydratase, encoded by gene *ilvA*, catalyzes the anaerobic formation of α -ketobutyrate from threonine in a two-step reaction. Since α -ketobutyrate is the initial substrate for ncBCAA biosynthesis, *ilvA* inactivation would then result in a reduced ncBCAA production. The *leu* operon, encoding for the leucine biosynthetic enzymes (*leuA*, *leuB*, *leuC* and *leuD*), also plays an important role in the ncBCAA biosynthesis since it catalyzes a number of enzymatic reactions (the so-called keto acid chain elongation process) from pyruvate to α -ketovalerate and α -ketocaproate, which are the precursors of norvaline and norleucine, respectively. Hence, inactivation of one or more genes of the *leu* operon would also decrease ncBCAA production. Fenton *et al.* (1994) demonstrated that by using *E. coli* strain CV512, which harbors the disrupting *leuA371* mutation, for recombinant protein production, no norleucine was detected. Bogosian *et al.* (1989) also reported that *E. coli* strain LBB24, harboring a deletion of the entire *leu* operon, could significantly reduce both norleucine and norvaline synthesis. However, both described strains required leucine supplementation to grow. Moreover, Bogosian *et al.* (1989) generated *E. coli* strain LBB254, which contains an *ilvA* disruptive mutation. Strain LBB254 was not able to grow without isoleucine supplementation. However, after induction of recombinant protein expression the effect of *ilvA* mutation was suppressed since strain was able to grow on minimal medium lacking isoleucine. This is explained because the amino acid sequence of the expressed recombinant protein (bovine somatotropin) has high leucine content (15 %) if compared with an average *E. coli* protein (9 %) so that after induction of protein expression leucine is depleted and, in consequence, transcription of the *leu* operon is activated, hence recovering formation of α -ketobutyrate, leucine, isoleucine and ncBCAA. In order to avoid the suppression effect reported in strain LBB254, Bogosian *et al.* (1989) generated *E. coli* strain LBB259, which lacks both *ilvA* and *leu* operon. This strain required leucine and isoleucine for growth. Purified bovine somatotropin expressed by strain LBB254 showed no norleucine incorporation.

Optimization of operational conditions and nutrient composition during fermentation can also lead to reduction of ncBCAA production. Ni *et al.* (2015) demonstrated that by using the Korz-based chemically defined medium in combination with a DO-stat fed-batch control a significant decrease of norvaline and norleucine concentrations was achieved without the need to supplement exogenous canonical amino acids.

Biermann *et al.*, 2013 demonstrated that under oxygen limitation and high glucose concentration conditions, media supplementation with trace elements molybdenum, selenium and nickel reduces norvaline and norleucine accumulation. These mentioned elements act as cofactors for the formate hydrogenlyase complex (FHL). FHL is a membrane-associated metalloprotein complex including a cytoplasmic molybdenum- and selenium-dependent formate dehydrogenase H (FdhF) and a transmembrane hydrogenase 3 complex consisting of 6 subunits: the [NiFe] hydrogenase component HycE, whose active site contains nickel and iron, the iron-sulfur proteins HycB, HycF and HycG and the integral membrane proteins HycC and HycD. FHL catalyzes formate oxidation to form CO₂ and H₂ under anaerobic conditions. According to Soini *et al.* (2008, 1) both norvaline and norleucine accumulate in *E. coli* cultivations under oxygen limitation and glucose excess. Under those conditions, glucose catabolism is enhanced and an intracellular accumulation of pyruvate occurs. This pyruvate excess leads, in turn, to a higher formation of α -ketobutyrate and, in consequence, of ncBCAA. A scenario triggering even higher ncBCAA formation under aforementioned conditions happens when inducing expression of leucine-rich recombinant proteins since leucine depletion would de-regulate *leu* operon thus increasing further the metabolic flux from pyruvate to ncBCAAs. FHL, together with lactate dehydrogenase (LDH) and pyruvate formate lyase (PFL), plays a key role in

the anaerobic metabolism of pyruvate. When using molybdenum, selenium and nickel, functionality of FHL is enhanced and part of the pyruvate excess generated can then be anaerobically metabolized to the detriment of the metabolic pathway catalyzed by the *leu* operon, which ends in ncBCAA biosynthesis. Both Biermann *et al.* (2013) and Soini *et al.* (2008, II) reported a reduction of formate accumulation in *E. coli* cultivations after supplementation of the aforementioned trace elements, which is directly related to a higher FHL functionality.

Moreover, Biermann *et al.*, 2013 reported that the degree of norleucine and norvaline biosynthesis strongly relies on the *E. coli* strain selected as recombinant protein expression system since strain *E. coli* BL21(DE3) reported a significant low accumulation of ncBCAA with respect to strain *E. coli* K-12. As opposed to strain *E. coli* BL21(DE3), genome sequencing and annotation studies revealed that a two-base insertion event between base pairs 1250 and 1253 is present in the coding sequence of gene *ilvG* in *E. coli* K-12 strains. This insertion causes a shift of the reading frame and, as a consequence, a stop codon is formed, resulting in a premature termination of *ilvG* gene expression (Yoon *et al.*, 2012; Parekh and Hatfield, 1997; Lawther *et al.* 1981). AHAS II is involved in biosynthesis of isoleucine starting from α -ketobutyrate and, in contrast to AHAS I and AHAS III, AHAS II is resistant to valine-mediated feedback inhibition. The absence of AHAS II activity in *E. coli* K-12 strains leads to valine toxicity: AHAS I and AHAS III are inhibited in the presence of valine and, since AHAS II cannot be properly synthesized, a drastic reduction of leucine and isoleucine biosynthesis occurs. Under these conditions, growth behavior is jeopardized (Anderson *et al.* 2001, Biryukova *et al.* 2010) and accumulation of α -ketobutyrate is enhanced, thus resulting in higher ncBCAA production (Soini *et al.*, 2008, I; Sycheva *et al.*, 2007).

Further strategies aiming reduction of ncBCAA mis-incorporation, which are not yet available in the state of the art, might be considered: (i) substitution of codons reporting higher ncBCAA mis-incorporation from the coding region of a gene of interest, (ii) protein engineering to increase specificity of biosynthetic enzymes for certain α -keto acids, (iii) protein engineering to increase specificity of aminoacyl tRNA synthetases for the canonical amino acids, and (iv) genetic regulation of target genes involved in the BCAA biosynthetic pathway.

Cvetesic *et al.* (2016) reported that norvaline is preferentially mis-incorporated at CTG codons, which represent 47 % of the total leucine codons available in an *E. coli* cell. They also demonstrated that leucine codons TTA, TTG and CTT show the most significant reduction of norvaline incorporation. Two options were hypothesized in order to explain that observations: (i) different aminoacylation kinetics for each tRNA^{leu} or (ii) different interaction of each Nva-tRNA^{leu} with the translation machinery. Accordingly, a substitution of CTG codons from the coding region of the gene of interest to others showing less mistranslation degree might lead to a reduction of norvaline mis-incorporation in the recombinant protein. Nevertheless, this might also in parallel decrease recombinant protein productivity since half of the total tRNA^{leu} present in *E. coli* contain anticodon CAG (Dong *et al.*, 1996), which might negatively affect translation rate. As for leucine, isoleucine can be encoded by alternative codons: ATT, ATC and ATA. However, no literature is available showing preference of β -methylnorleucine mis-incorporation for one of the mentioned codons. Additionally, Veeravalli *et al.* (2015) demonstrated that, contrary to previous reports (Bogosian *et al.*, 1989), norleucine mis-incorporation is not random since, at certain amino acid positions, higher levels of norleucine are reported. However, this observation cannot be explained by alternative codon usage since only codon ATG is available for methionine.

α -isopropylmalate synthase (α -IPMS) participates in the so called α -ketoacid chain elongation pathway. Although the main substrate for α -IPMS is α -ketoisovalerate (α -KIV), the enzyme also shows certain affinity towards other α -ketoacids such as pyruvate, α -ketobutyrate (α -KB) or α -ketovalerate (α -KV) (Table 5). This enzymatic promiscuity facilitates biosynthesis of ncBCAA. Similar to α -IPMS, α -acetohydroxyacid synthase (AHAS) has broad substrate specificities (Table 6). Depending on the isoenzyme (AHAS I, II or III), AHAS can transform 2 pyruvate molecules into α -acetolactate and/or pyruvate and α -ketobutyrate into α -acetohydroxybutyrate. However, AHAS can also convert pyruvate and α -ketovalerate into α -aceto- α -hydroxyvalerate, which is a precursor of β -methylnorleucine (Muramatsu *et al.*, 2003; Gollop *et al.* 1989). Accordingly, protein engineering approaches leading to reduce promiscuity of the leucine and valine/isoleucine biosynthetic enzymes would be considered as strategy to decrease chain elongation of α -ketoacids leading to biosynthesis of ncBCAA.

Moreover, as discussed by Veeravalli *et al.* (2015), usage of aminoacyl tRNA synthetases from other organisms which cannot use ncBCAA as substrate or engineering of aminoacyl tRNA synthetases to increase their specificity against cBCAA could be promising approaches to prevent mis-incorporation of ncBCAA into recombinant proteins manufactured in *E. coli*. This option might be feasible in the near future since there is increasing knowledge of the aminoacyl tRNA synthetase structure and its editing mechanisms.

Another approach which has not yet been explored is the expression regulation of genes involved in the BCAA biosynthetic pathway in order to minimize ncBCAA biosynthesis and this strategy is the focus of the current thesis.

2.4 Recombineering in *E. coli*

Recombineering refers to recombination-mediated genetic engineering. Recombineering comprises molecular techniques which enable modification of a DNA target by homologous recombination. Homologous recombination is defined as the pairing of homologous sequences (regions of identical sequence) between target DNA and linear donor DNA so that the DNA sequence located between homologous sequences is replaced, resulting in recombinant DNA. Linear DNA that is either single- (ss-) or double-stranded (-ds) can be used as donor DNA for recombineering. ssDNA are normally synthetic oligonucleotides while dsDNA can be easily obtained by PCR. A variety of genetic modifications can be carried out depending on design of the linear donor DNA: gene replacements, deletions and point mutations. This study focuses on in-frame gene knock-outs (Court *et al.*, 2002).

Homologous recombination naturally occurs in *E. coli* and it plays an important role in repairing DNA damage during replication. Homologous recombination is mediated by the orchestrated action of proteins RecA, RecBCD and RecF, encoded by genes *recA*, *recBCD* and *recF*, respectively. RecA is recombinase, RecBCD a helicase/nuclease complex and RecF a DNA replication and repair protein. RecA binds to single-strand DNA sections forming a RecA-DNA complex able to pair with other homologous DNA regions. When homology is encountered, RecA triggers a DNA exchange between both strands. RecBCD plays a major role in initiating recombination at double-strand DNA boundaries by creating 3'-single-strand overhangs upon degradation of 5'-ends. RecBCD also assists RecA in binding to ssDNA regions. Similar to RecBCD, RecF starts recombination events at dsDNA terminus.

However, RecF recombination activity is around 100-fold lower than RecBCD (summarized by Court *et al.*, 2002).

Recombineering demands transformation of linear donor DNA into *E. coli* hosting efficient homologous recombination machinery that enables exchange of DNA target with donor DNA. The bacterial-encoded recombination system has been proven to be inefficient for recombination of exogenous linear DNA. Moreover, transformability of *E. coli* is low due to degradation of linear DNA by intracellular exonucleases. Thus, in order to enhance linear DNA recombination efficiency in *E. coli*, alternative recombination systems have been explored. The bacteriophage λ -encoded Red recombination system has been proven to be at least 50-fold more efficient than previous systems (Murphy, 1998). The λ Red recombination system comprises Exo, Beta and Gam proteins, encoded by genes *exo*, *bet* and *gam*, respectively. While Gam inhibits the bacterial endogenous RecBCD nuclease activity, Exo and Beta triggers homologous recombination. Hence, linear donor DNA is not degraded and can recombine with the bacterial genome. Exo has 5'→3'-nuclease activity. It degrades 5'-ends of the linear donor DNA and creates 3'-ssDNA overhangs. Beta can then bind to the generated ssDNA overhangs and triggers pairing of donor DNA with the homologous DNA region present in the *E. coli* genome. Datsenko and Wanner (2000) developed a simple and highly efficient λ Red-based recombination method to knock-out single genes from the *E. coli* chromosome by using dsDNA PCR-products containing 36 bp homology arms as linear donor DNA. This methodology was later employed to generate the so-called KEIO collection, which comprises a palette of single gene knock-out *E. coli* strains (Baba *et al.*, 2006).

λ Red recombination system can be expressed from a defective prophage integrated in the *E. coli* chromosome or from a helper plasmid. In this thesis, the low-copy helper plasmid pKD46 is employed. pKD46 carries genes *exo*, *bet* and *gam* under the control of an arabinose-inducible promoter. Plasmid contains a pSC101-based origin of replication (*ori*), allowing a low-copy number. This *ori* is temperature-sensitive so that plasmid curation can be achieved by cultivating hosting *E. coli* at 42 °C. Moreover, pKD46 contains gene *araC*, which encodes the repressor AraC protein, thus preventing basal expression of *P_{BAD}* promoter and allowing tight expression regulation of λ Red genes. *E. coli* K-12 BW25113 was selected in this study for recombineering purposes. Since *E. coli* K-12 BW25113 is deficient in arabinose catabolizing enzymes – it contains a deletion of the *araBAD* operon: $\Delta(araD-araB)567$ – arabinose can be successfully employed as inducer of λ Red genes hosted in plasmid pKD46 (Datsenko and Wanner 2010).

In this study, the FRT/FLP homologous recombination system was employed to knock-out operons *ilvH* and *ilvBN* as well as genes *thrA*, *leuA*, *ilvC* and *ilvA* from *E. coli* K-12 BW25113 genome.

2.4.1 Homologous recombination based on FRT/FLP system

The FRT/FLP system originates from *Saccharomyces cerevisiae* and consists of a site-specific flipase recombinase (FLP) and the FLP recombinase recognition sites (FRT). FLP triggers site-specific recombination between FRT sites. This strategy is mainly employed to generate in-frame chromosomal gene deletions. In this approach a linear dsDNA fragment containing at both ends around 36 bp homologous overlaps to the target gene is generated by PCR. Moreover, this dsDNA fragment carries an antibiotic resistance cassette flanked by two FRT sites. In the current study, available plasmids pKD3 or pKD4 were used as templates to generate linear donor dsDNA by PCR.

While plasmid pKD3 contains a chloramphenicol-resistance gene flanked by FRT sites and two priming sites, plasmid pKD4 shares the same features than pKD3, but contains a kanamycin-resistance gene. PCR primers contain 20 bp homology to the priming sites of the template plasmids followed by 36 bp ends homologous to the flanking genomic region of the target gene (Datsenko and Wanner, 2000).

At first, the generated dsDNA fragment is transformed into *E. coli* and replaces target gene by λ Red-mediated homologous recombination. λ Red recombination system is normally expressed in a helper plasmid such as pKD46 (Datsenko and Wanner, 2000). Potential recombinant clones are then selected for antibiotic resistance. Secondly, antibiotic resistance marker is excised from the genome by FLP-mediated site-specific recombination between the two FRT sites. FLP is also normally provided by a helper plasmid such as pCP20. FLP synthesis in plasmid pCP20 is under control of a temperature-sensitive λ repressor and induced at 30 °C (Cherepanov and Wackernagel 1995, Datsenko and Wanner 2000). Similar to plasmid pKD46, pCP20 contains a temperature-sensitive pSC101-based origin of replication, allowing easy plasmid curation at 42 °C (Datsenko and Wanner, 2000). Finally, removal of the antibiotic resistance marker leaves in the genome a 103 bp-scar sequence containing an active FRT site (Figure 7). Accordingly, the FRT genetic scar is considered suboptimal when several genes need to be altered in the same clone, since with every genetic modification an additional FRT site would be added to the genome. Since those FRT sites are also active, the risk of genomic misintegration or even deletions of big chromosomal regions increases the more sequential modifications are performed. Datsenko and Wanner (2000) recommended using this strategy for single gene deletions. Reitz (2011) could perform up to 3 knock-outs in a single clone with this strategy.

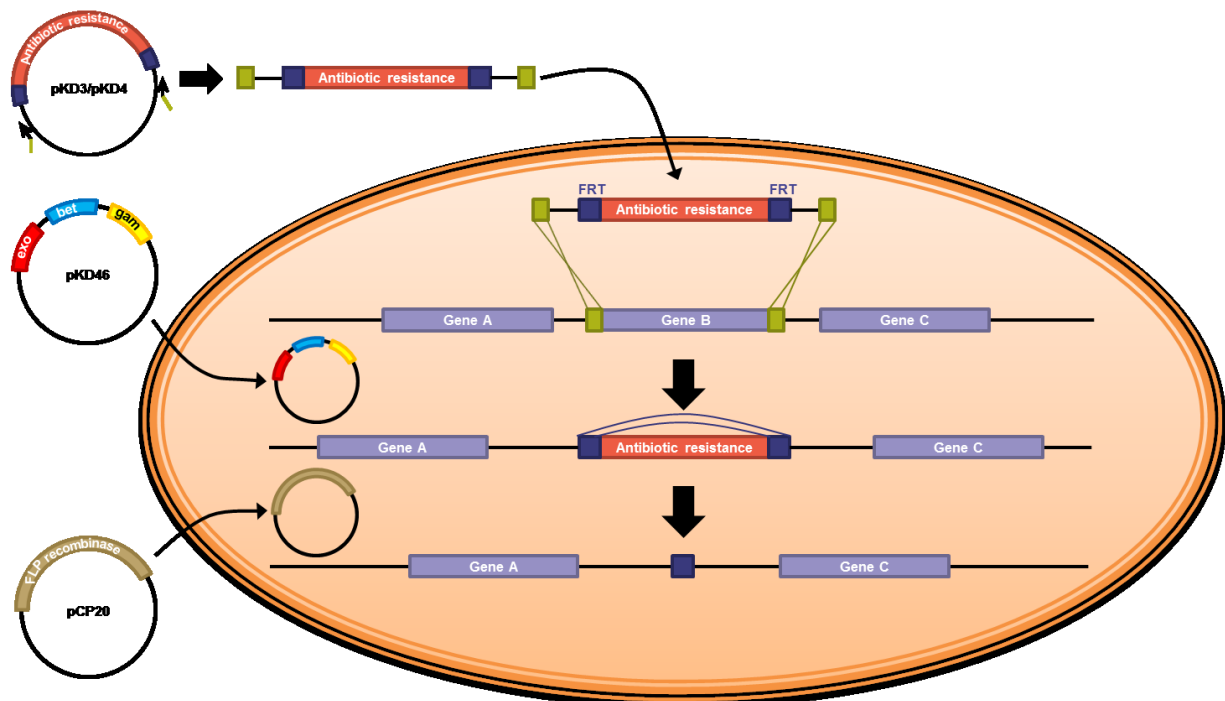


Figure 7. Schematic representation of the λ Red-mediated gene replacement with selection marker elimination by site-specific recombination. See text for description. Adapted from Madyagol *et al.* (2011).

2.5 Tunable promoter systems in *E. coli* metabolic engineering

Various tunable promoter systems have been used for high-level expression of recombinant proteins in *E. coli*. Nevertheless, an excessive, non-physiological, presence of recombinant protein can lead to non-desirable effects, such as the generation of inclusion bodies and cell metabolic burden. Tunable promoter systems are also convenient for metabolic engineering applications where protein expression at physiological levels is usually required. This section offers a summary of the molecular mechanisms and applications of four of the most frequently used tunable promoter systems in metabolic engineering of *E. coli* (*araBAD*, *rhaBAD*, *lac* and *xyIS/Pm*) (Terpe, 2006), as well as inherent advantages and drawbacks.

2.5.1 Introductory considerations

In order to achieve high-level expression of recombinant protein in *E. coli*, genes encoding the recombinant protein are usually cloned under the control of inducible promoters on high-copy plasmids. However, an excessive recombinant protein expression can lead to non-desirable effects, such as unwanted aggregation and cell metabolic burden. As a result of the forced metabolic burden, cell physiology might be noticeably affected, being specific growth rate and protein yield the main altered parameters. In these conditions, plasmid free cells tend to grow faster and plasmid stability is harmed. Metabolic burden effects are more severe the bigger the plasmid is and the more plasmids are present in the cell (Glick, 1995; Wu *et al.*, 2016). Therefore, the use of tunable promoter systems in medium- or low-copy plasmids would reduce the metabolic burden imposed to the cells, thus minimizing the problems mentioned before (Keasling, 1999; Jones *et al.*, 2000).

Metabolic engineering is the application of genetic engineering in order to refashion the metabolism of a certain organism to improve cellular properties and to increase the production of a target metabolite. Metabolic engineering includes the optimization of existing metabolic pathways as well as the introduction of new ones. The field of metabolic engineering appeared in 1991 (Bailey, 1991). Since then, plenty of literature is available, demonstrating success of metabolic engineering approaches in *E. coli*. The commercialization of the chemical building blocks 1,3-propanediol and 1,4-butanediol manufactured by engineered *E. coli* are two examples of success (Nielsen *et al.*, 2016). Most recent studies achieved the production of L-methionine (Li *et al.*, 2017), monolignol (Chen *et al.*, 2017), astragalin (Pei *et al.*, 2016), a plant-base anti-inflammatory agent (Ahmadi *et al.*, 2016), L-malate (Dong *et al.*, 2016) and ectoine (Ning *et al.*, 2016) through metabolic engineering. For successful metabolic engineering, genetic tools employed should fulfill a number of requirements: (i) genetic construct carrying the genes should be stable in the host over generations and should not impose cell metabolic burden, (ii) promoter triggering expression of the target genes should be tightly regulated, being expression totally repressed in absence of inducer and activated in a linear dose dependent manner in its presence, and (iii) expression of target genes should be comparable in all cells of the culture at any time.

Hence, for metabolic engineering applications, where protein expression at physiological levels is usually required, use of low-copy number plasmids containing target genes is encouraged for proof-of-concept applications (Jones and Keasling, 1998). Alternatively, integration of the target genes into the chromosome would improve gene stability and minimize metabolic burden. As opposed to

plasmid systems, chromosomal integration does not require the use of selective pressure for maintenance of genetic material. However, expression of genomically integrated genes is dramatically affected by their position relative to the chromosomal origin of replication (Sousa *et al.*, 1997; Couturier and Rocha, 2006).

For most metabolic engineering purposes, promoters allowing positive regulation of gene expression (tunable promoters) in presence of inducer are favored. However, it is important to consider that inducers are differently transported into the cell and that some of them are more predisposed to be metabolized. An ideal inducer would be the one entering the cell by passive diffusion and not being metabolized by the cell machinery, which would allow homogenous and persistent protein expression in all cells in cultivation, thus avoiding the all-or-none induction effect. As aforementioned, in the absence of inducer, tunable promoter should be totally repressed while promoter expression should vary in a positive linear dose dependent manner in its presence. Inducer should also be inexpensive and non-toxic.

This section offers a summary of the molecular mechanisms and applications of four of the most frequently used tunable promoter systems in metabolic engineering (*araBAD*, *rhaBAD*, *lac* and *XylS/Pm*), as well as inherent advantages and drawbacks and possible improvements.

2.5.2 *araC-P_{BAD}* promoter system

The arabinose-inducible *araBAD* promoter (P_{BAD}) together with its regulator protein AraC has become a popular tunable expression system for high-level recombinant protein production as well as for metabolic engineering purposes since expression can be tuned over a broad range of arabinose concentrations (Guzman *et al.*, 1995).

In the absence of arabinose, AraC coils DNA by binding to *araI*₁ and *araO*₂ thus blocking RNA polymerase to bind to P_C and P_{BAD} promoters and cyclic AMP receptor protein (CRP) to bind to its respective binding region besides *araI*₁. When arabinose is present in the medium, it is transported into the cells by the low-affinity high-capacity AraE and the high-affinity low-capacity AraFGH arabinose transporter systems. If intracellular arabinose concentration surpasses a certain threshold concentration, it binds to and activates AraC, which in turn undoes the DNA loop, binding to *araI*₂ instead of *araO*₂. This conformation allows binding of CRP and interaction of RNA polymerase with the promoter regions, thus stimulating expression from promoters P_{BAD} and P_C . Promoters P_E and P_{FGH} are also induced when arabinose is present. However, when AraC binds to *araO*_{1L} and *araO*_{1R}, promoter P_C is repressed. *AraBAD* are enzymes involved in arabinose catabolism. Furthermore *araC-P_{BAD}* promoter system is also regulated by catabolite repression (Schleif, 2000; Megerle *et al.*, 2008).

Its inherent characteristics make *araC-P_{BAD}* promoter system especially attractive since basal levels of expression are particularly low (tight regulation) and response to arabinose is very fast. However, some literature shows that the expression triggered by this promoter is not homogenous at a cell population level due to the all-or-none induction mechanism (Siegele und Hu, 1997; Carrier and Keasling, 1999; Giacalone *et al.*, 2006; Brautaset *et al.*, 2009; Balzer *et al.*, 2013; Binder *et al.*, 2016), being then problematic when used for metabolic engineering purposes. As mentioned before, arabinose enters the cell via protein transport systems AraE and AraFGH. However, distribution of arabinose transporters in a cell population is not homogeneous due to dilution by cell division and

the arabinose-dependent expression of genes *araE* and *araFGH* that causes a stochastic background expression of promoters P_E and P_{FGH} , possibly as a result of a stochastic change in conformation of AraC while interacting with the promoter region. Consequently, two fractions of cells are present. On the one hand, the fraction of cells containing enough transporters to accumulate intracellular arabinose beyond the threshold concentration induces the synthesis of new transporters that uptake more arabinose, thus amplifying induction. On the other hand, the fraction of cells bearing insufficient transporters does not accumulate intracellular arabinose beyond the threshold concentration necessary for induction and hence, cells are not induced. The likelihood that a cell is induced increases the higher is the arabinose concentration present in the medium (Siegele und Hu, 1997; Morgan-Kiss *et al.*, 2002; Széliová *et al.*, 2016).

Numerous strategies have been proven to be satisfactory for elimination of the all-or-none induction effect inherent to the *araC*- P_{BAD} promoter system, allowing then homogenous expression at population level: (i) deletion of the *araFGH* operon and substitution of the P_E promoter controlling expression of the AraE transporter by an IPTG inducible promoter (Khlebnikov *et al.*, 2000, 2002), (ii) deletion of the *araBAD* and *araFGH* operons and substitution of the P_E promoter controlling expression of the AraE transporter by a constitutive promoter (Khlebnikov *et al.*, 2001), (iii) deletion of the *araBAD* operon and both AraE and AraFGH arabinose transport systems and use of a lactose transporter (LacY) of relaxed specificity as an L-arabinose transporter (Morgan-Kiss *et al.*, 2002), (iv) deletion of the *araBAD* operon and substitution of the P_E promoter controlling expression of the AraE transporter by the stronger P_{BAD} promoter (Széliová *et al.*, 2016) and (v) deletion of the *araBAD* operon and use of photocaged arabinose inducers (Binder *et al.*, 2016).

Strategies (i) and (ii) focus on uncoupling expression of the native low affinity-high capacity AraE transporter from arabinose-dependent regulation. Strategy (iii) completely substitutes the native arabinose transport systems by an analog transporter, whose expression is uncoupled from arabinose-dependent regulation. Strategy (iv) allows a general increase of the basal amount of native arabinose transporters present in the cells, hence decreasing the arabinose concentration threshold necessary to trigger induction, which in turn results in a more homogenous protein expression at population level. However, expression of AraE is still dependent on arabinose. Strategy (v) achieves population homogeneity as photocaged arabinose might enter the cell by passive diffusion, not depending on transport systems.

2.5.3 *rhaBAD* promoter system

The rhamnose-inducible *rhaBAD* promoter together with its regulator proteins RhaR and RhaS has demonstrated to be tunable over a wide range of rhamnose concentrations (Giacalone *et al.*, 2006).

The *rhaBAD* promoter shares lots of features with the *araBAD* promoter. Rhamnose is transported into the cells by RhaT and it is then metabolized by RhaBAD. When rhamnose is present in the intracellular medium, it binds to the regulator protein RhaR, which in turn activates transcription of *rhaR* and *rhaS* promoters. RhaR is constitutively expressed at low levels. The regulator protein RhaS also binds to rhamnose, hence activating transcription from *rhaT* and *rhaBAD* operon. Transcription of *rhaBAD* promoter is also regulated by catabolite repression, since it requires the binding of CRP (Egan and Schleif, 1993; Via *et al.*, 1996).

There are also evidences that L-rhamnose promoter exhibits the 'all-or-none' response in *E. coli* (Ozbudak *et al.*, 2004; Giacalone *et al.*, 2006; Brautaset *et al.*, 2009; Afroz *et al.*, 2014; Kelly *et al.*, 2016) and other microorganisms (Cardona und Valvano, 2005). L-rhamnose concentration affects the relative fraction of cells in the fully induced or uninduced states (Afroz *et al.*, 2014; Ozbudak *et al.*, 2004). Similarly to L-arabinose, L-rhamnose catabolism involves the inducible expression of the transport system RhaT by rhamnose, so this might explain how *rhaBAD* promoter induction could develop bimodality by this mechanism (Kelly *et al.*, 2016). Despite that, no literature is available for *rhaBAD* promoter concerning possible strategies minimizing the all-or-none induction effect but a feasible option would be the use of strains defective in rhamnose-catabolizing enzymes.

In order to successfully use the *rhaBAD* promoter for metabolic engineering purposes, it might be convenient to use genetically engineered *E. coli* strains where the *rhaBAD* operon is depleted, hence avoiding metabolization of rhamnose. In addition, similarly to the *araBAD* promoter, the reported all-or-none induction mechanism may be solved by uncoupling expression of the native RhaT transporter from rhamnose-dependent regulation; i.e. substitution of the native promoter controlling expression of the RhaT transporter by a constitutive promoter.

2.5.4 *lac* promoter system

The lactose operon is formed by genes *lacZ*, *lacY* and *lacA*. Lactose is transported into the cells by a lactose permease, encoded by *lacY*. Lactose is then metabolized into glucose and galactose by β -galactosidase, which is encoded by *lacZ*. The gene *lacA* encodes a galactoside O-acetyltransferase which might have an important role in cellular detoxification by acetylating non-metabolizable pyranosides, thus reducing rate of retransport into the cell (Roderick, 2005; Marbach and Bettenbrock, 2012). *lacA* does not seem to be involved in lactose metabolism. The lactose operon is regulated by the LacI repressor, which is encoded by the gene *lacI*. LacI binds to the operator region of the operon, thus inhibiting transcription of genes *lacZ*, *lacY* and *lacA*. However, when lactose is present in the medium, allolactose can bind to the LacI repressor and triggers the release of the repressor from the operator region, thus activating transcription of genes *lacZ*, *lacY* and *lacA*. However, binding of the cAMP-bound catabolite activator protein (CAP) is also required for transcription activation, being then *lac* operon subjected to catabolite repression (Busby and Ebright, 2001; Görke and Stülke, 2008).

Allolactose is the natural inducer of the *lac* operon. However, it can be metabolized by β -galactosidase and this is not advantageous for metabolic engineering applications since a constant induction of expression of target genes is preferable. However, artificial structural analogs of allolactose such as TMG (methyl-1-thio- β -d-galactopyranoside) and IPTG (isopropyl β -d-1-thiogalactopyranoside), that cannot be metabolized, have been developed (Daber *et al.*, 2007; Marbach and Bettenbrock, 2012). Similar to lactose, both artificial inducers are also transported into the cell by LacY at low inducer concentrations. Nevertheless, at high inducer concentrations an alternative transport independent on LacY is possible as well. As aforementioned, LacA can acetylate TMG and IPTG, thus becoming inactive as inducer molecules (Marbach and Bettenbrock, 2012).

Numerous variants of the original *E. coli lac* promoter have been designed for molecular biology applications: lacUV5, P_{tacl}, P_{trc} and T7lac. The lacUV5 promoter contains 2 point mutations in the -10 promoter region and an additional one at -66, thus differing in only 1 nucleotide from the consensus

sequence (Hirschel *et al.*, 1980). As a result, the lacUV5 promoter features a higher affinity for *E. coli* RNA polymerase than the wild type *lac* promoter. The *lac* operon driven by the UV5 promoter achieves about 50 fold higher expression than the wild type version (Pribnow, 1975) and does not require the presence of cAMP-CAP, that is, is insensitive to catabolite repression (Arditti *et al.*, 1973; Silverstone *et al.*, 1970). P_{tac} is a hybrid promoter where -35 region is derived from the *trp* promoter while -10 region comes from the lac UV5 promoter, being about 10 times more efficient than the parental lac UV5 promoter (de Boer *et al.*, 1983). P_{trc} is a variant of P_{tac} where one additional base pair was inserted between -35 and -10 promoter regions, having about 90% the activity of P_{tac} . The T7lac promoter is a variant of the UV5 promoter where the gene encoding T7 RNA polymerase is under the control of the lacUV5 promoter genomically integrated while the gene of interest is located in a plasmid under the control of a T7 promoter. This system offers a very tight regulation and achieves about 4 fold higher expression than the P_{trc} promoter (Tegel *et al.*, 2011).

lac-based promoters are specially leaky, that means, expression is also triggered in a basal level even when no inducer is present. This is because LacI repressor does not tightly bind to the operator and can dissociate. Hence, when using *lac*-based promoters in cloning vectors is important to take into consideration that enough LacI repressor is present in order to avoid leaky expression of the promoter. The number of LacI molecules necessary to repress *lac* promoter mainly depends on the number of operators regulating the promoter, the affinity of the repressor to the operator and the copy number of the plasmid bearing the *lac* promoter (Penumetcha *et al.*, 2010). Several strategies have been developed in order to improve the regulation tightness of *lac*-based promoters: use of low copy plasmids for recombinant protein expression, overexpression of LacI repressor by using stronger *lacI* promoter variants and use of LacI variants with higher affinity towards operator regions. Different variants of *lacI* promoter sequences leading to different expression levels of *lacI* have been reported in *E. coli* strains: *lacI*⁺ and *lacI*^q in Calos (1978), I-UJ177 in Calos and Miller (1980) and *lacI*^{q1} in Calos and Miller (1981). According to Glascock *et al.* (1998), an *E. coli* wild type cell (*lacI*⁺ variant) has about 10 LacI molecules. The *lacI*^q promoter version shows 10-fold expression enhancement of lac repressor if compared with the *lacI*⁺ variant (100 *lacI* molecules) while expression by *lacI*^{q1} is 17-fold stronger than that of *lacI*^q and 170-fold stronger than that of *lacI*⁺ (1700 *lacI* molecules). In addition, numerous *lacI* gene mutants conferring a higher binding affinity of the repressor for the operator region have been utilized: LacI-I12, LacI-X86 and the double mutant LacI-I12/X86. LacI-I12 and LacI-X86 confers a 50-100 fold improvement of binding affinity if compared with the wild type LacI while the LacI-I12/X86 variant causes a 10000 fold increase (Penumetcha *et al.*, 2010).

Similarly to the *araBAD* promoter, the all-or-none induction effect was first observed for the *lac* operon by Novick and Weiner (1957). LacY catalyses the uptake of IPTG and TMG, which in turn induces further expression of LacY, thus leading to a positive feedback loop that is responsible of the bimodal behavior of gene expression at population level (Ozbudak *et al.*, 2004). However, Marbach and Bettenbrock (2012) and Binder *et al.* (2014) demonstrated that a *lacY* deletion results in a homogeneous response to induction at population level. However, when using this mutant, a 20-fold higher TMG and 10-fold higher IPTG concentration was necessary to achieve similar induction levels than the wild type strain (Marbach and Bettenbrock, 2012).

2.5.5 Pm/XylS promoter system

It has recently become a popular promoter system. Plenty of literature is available proving the efficiency of the Pm/XylS expression system (reviewed by Brautaset *et al.*, 2009).

XylS is constitutively transcribed by its own promoter Ps2, and the gene of interest is placed under the control of Pm promoter. The inducer molecule (benzoic acid derivatives) enter the cells via passive diffusion and bind to XylS, which then activates transcription of the gene of interest from Pm promoter (Brautaset *et al.*, 2009).

Unlike the previously mentioned lactose, arabinose and rhamnose-based expression systems, benzoate inducers for the activation of Pm/XylS systems don't rely on active transport systems but enter the cells via passive diffusion (Brautaset *et al.*, 2009; Binder *et al.*, 2016). Hence, homogeneous expression at the cell population level is reported since no regulated transported system is present (Balzer *et al.*, 2013; Binder *et al.*, 2016) and no previous strain genetic engineering is needed.

Moreover this system is flexible, since it does not require specific features of the host, in contrast to most other systems. Hence, the performance of this expression system is not strain dependent in *E.coli*. Induction is carried out with benzoic acids, which are very cheap and non-toxic. Conventional induction with m-toluate especially leads to a minimum growth impairment and homogenous population (Binder *et al.*, 2016). However, the system leads to slight protein expression also in the absence of inducer and that expression level is especially high for the Pm variant ML1-17 (Balzer *et al.*, 2013). According to Binder *et al.* (2016), benzoate induction systems worked better at 30°C.

Pm/XylS expression system is also commercially available (Website 1). This expression system is considered optimal for metabolic engineering when used in conjunction with minimal replicons such as RK2. One of these vectors, pJB658, has proven useful for tightly regulated recombinant gene expression in several gram-negative species (Blatny *et al.*, 1997; Brautaset *et al.*, 2000; Winther-Larsen *et al.*, 2000 and EP2078076 B1; Sletta *et al.*, 2004).

2.5.6 Summary

To sum up, *araBAD*, *rhaBAD* and *lac*-based promoters are not optimal since they exhibit an 'all-or-none' expression response at cell population level. However, for *araBAD* and *lac* promoters, strategies minimizing that negative effect are available. Hence, in order to successfully implement these promoters for metabolic engineering strategies, previous strain engineering is required. As opposed to aforementioned promoters, Pm/XylS promoter shows homogeneous expression response at population level thanks to diffusion-based transport into the cell so that no previous strain engineering is necessary. Moreover, Pm/XylS promoter seems promising since expression level is tunable and the inducer is non-toxic and inexpensive.

3. Research Hypotheses and Aim of the Project

The incorporation of non-canonical branched chain amino acids (ncBCAA) such as norleucine, norvaline and β -methylnorleucine into recombinant proteins during *E. coli* production processes has become a crucial matter of contention in the pharmaceutical industry, since such mis-incorporation can lead to the production of altered proteins, having non optimal characteristics. Although several strategies have been developed in order to reduce ncBCAA mis-incorporation in *E. coli*, a number of limitations impede them to be effectively applied in large-scale recombinant protein production processes. This thesis focuses on another approach which has not yet been explored and that might contribute to close the scientific gap.

The main objective of this thesis is the generation of new genetically engineered *E. coli* strains allowing the production of recombinant proteins with a considerably reduced content of ncBCAA. Since ncBCAA biosynthesis is controlled by enzymes encoded by genes involved in the BCAA biosynthetic pathway such as the *leu*- and *ilv*-encoded enzymes, it would be valuable to have a more comprehensive understanding of the effect of the expression regulation of the implicated genes into ncBCAA biosynthesis. Hence, the novel *E. coli* strains would enable exogenous tunable expression of target genes involved in the BCAA biosynthetic pathway so that the expression level of a certain target gene reporting a minimized ncBCAA biosynthesis and mis-incorporation could be identified. Screening of the engineered *E. coli* strains was performed in a mini-reactor platform by cultivating them in fed-batch mode under standard cultivation conditions. Expression levels of each target gene were determined by the amount of L-arabinose added in each experimental case. At the end of the cultivation, ncBCAA concentrations were determined in the inclusion body and intracellular soluble fraction so that for each tested target gene the expression level resulting in the further reduction of ncBCAA production could be identified.

However, since potential novel *E. coli* strains are aimed to be used in recombinant protein production processes taking place in large-scale reactors, it might be that data resulting from screening in mini-reactors under aforementioned cultivation conditions would not be conclusive since large-scale effects are not properly reproduced in that experimental design. As cultivation screenings cannot be directly performed at pilot or production scale due to capacity and economical reasons, application in the mini-reactor system of scale-down approaches simulating environmental perturbations occurring at large-scale would be suitable to obtain more realistic data. One of the main causes triggering ncBCAA biosynthesis in *E. coli* is the metabolic shift from oxidative respiration to overflow metabolism driven by inefficient mixing in large-scale reactors (Enfors *et al.*, 2001). Under these conditions, not only the mixed-acid fermentation products accumulate, but also pyruvate (Soini *et al.*, 2008, 1). Pyruvate excess present intracellularly increases the metabolic flux going to ncBCAA biosynthesis through the sequential keto acid chain elongation from pyruvate to α -ketocaproate over α -ketobutyrate and α -ketovalerate by the actuation of the *leu* operon-encoded enzymes (Apostol *et al.*, 1997). Hence, another objective of this thesis was to apply in the mini-reactor system cultivation conditions triggering ncBCAA biosynthesis, as it happens in industrial scale. A *bis dato* non-explored cultivation strategy combining pyruvate pulses and oxygen limitation was employed in this work in order to reproduce large-scale effects. This novel scale-down approach may have an advantage with respect to other strategies recently described in single compartment

reactors consisting on glucose pulses combined with oxygen limitation (Anane *et al.*, 2019) or oxygen down-shift combined with glucose excess (Soini *et al.*, 2008, I).

After screening of the engineered *E. coli* strains in the mini-reactor system under both standard cultivation conditions and conditions reproducing large-scale effects i. e. pyruvate pulsing and oxygen limitation, the most promising *E. coli* mutants were selected. The last objective of this thesis was to test those potential *E. coli* mutants expressing the optimal amount of the target gene in a 15L reactor under cultivation conditions subjected to pyruvate pulsing and oxygen limitation in order to verify its superior performance concerning recombinant protein impurity profile in comparison with the non-engineered *E. coli* strain.

Considering the aforementioned, the following research hypotheses have been contemplated to build up the basis of this current investigation:

- *Hypothesis 1. Down-regulation of single genes leuA, thrA or ilvA, up-regulation of single operons ilvBN, ilvIH or ilvGM as well as up-regulation of single gene ilvC would reduce ncBCAA biosynthesis and subsequent mis-incorporation into recombinant proteins expressed in E. coli.*

Taking into consideration the BCAA metabolic pathway, following strategies were pursued by modulating expression of target genes in *E. coli* in order to reduce ncBCAA biosynthesis: (i) limit conversion of pyruvate to α -ketobutyrate, (ii) limit transformation of threonine to α -ketobutyrate and (iii) limit conversion of α -ketobutyrate to α -ketovalerate. Strategy (i) might be achieved by down-regulating operon *leuABCD* but also by up-regulating operon *ilvBN*, strategy (ii) may be realised by down-regulating the *thr* genes as well as *ilvA*. Strategy (iii) could be accomplished by down-regulating operon *leuABCD* and up-regulating *ilvIH*, *ilvGM* or *ilvC*. According to this hypothesis novel *E. coli* strain mutants were genetically engineered so that the expression of single target genes (*leuA*, *thrA*, *ilvA*, *ilvC*, *ilvIH*, *ilvBN* and *ilvGM*) could be modulated in order to evaluate the effect of genetic modulation in ncBCAA biosynthesis.

The hypothesis was tested for each target gene by screening each engineered *E. coli* mutant expressing different levels of the target gene in a mini-reactor system. During the screening, the influence of a certain expression level of each target gene in ncBCAA biosynthesis and mis-incorporation into the recombinant protein was evaluated. Performance of the most promising *E. coli* strain mutants selected after screening was additionally verified in a 15L-reactor scale.

- *Hypothesis 2. Cultivation conditions subjected to pyruvate pulses and oxygen limitation represent a novel strategy reproducing large-scale effects, i.e. biosynthesis of ncBCAA and overflow metabolism by-products*

According to this scenario, pyruvate would enter the cell and directly accumulate there, hence triggering formate, acetate and ncBCAA biosynthesis.

The hypothesis was verified by comparing levels of formate, acetate and ncBCAA over cultivation time between the *E. coli* cultivation performed under standard conditions and the cultivation subjected to pyruvate pulses and oxygen limitation.

- *Hypothesis 3. Cultivation conditions subjected to pyruvate pulses and oxygen limitation represent a novel strategy reproducing large-scale effects, i.e. biosynthesis of ncBCAA and*

overflow metabolism by-products, faster than in scale-down cultivation strategies based on glucose excess.

When using glucose pulses, large-scale effects (i. e. increase of ncBCAA biosynthesis) can first be reported once glucose is converted to pyruvate and this starts accumulating in the cell. However, when using pyruvate pulses, pyruvate would directly enter the cell and would accumulate faster intracellularly, resulting in a rapid trigger of ncBCAA biosynthesis. This scale-down approach would then accelerate the velocity of the strain screening process.

The hypothesis was verified by comparing the time passed from the moment the perturbation was applied until a significant increase of ncBCAA concentrations was reported between the cultivation strategy used in this study (pyruvate pulses and oxygen limitation) and other published approaches based on glucose excess.

- *Hypothesis 4. Metabolic effects triggered by cultivation conditions subjected to pyruvate pulses and oxygen limitation translate first in the cytosol as ncBCAA biosynthesis and accumulation and then in the recombinant protein as ncBCAA mis-incorporation.*

NcBCAA are first produced by the BCAA biosynthetic pathway in the cytosol and accumulate intracellularly. After a certain accumulation, ncBCAA can then be mis-incorporated into the expressed recombinant protein by the translation machinery.

The hypothesis was tested by comparing the time passed from the moment the perturbation was applied until a significant increase of ncBCAA concentrations was reported in both inclusion body and intracellular soluble protein fraction.

- *Hypothesis 5. The probability that a certain cAA is exchanged by the respective ncBCAA analog depends on the number of individual positions of such cAA in the sequence of the expressed recombinant protein.*

The amount of leucine residues in the recombinant mini-proinsulin is higher compared to methionine or isoleucine. Hence, recombinant protein would report higher mis-incorporation of norvaline, compared to norleucine and β -methylnorleucine.

The hypothesis was verified by comparing the concentrations of each ncBCAA that were measured in the inclusion body fraction.

- *Hypothesis 6. Induction of recombinant mini-proinsulin expression has a positive effect on ncBCAA biosynthesis and mis-incorporation.*

Overexpression of leucine-rich recombinant proteins cause depletion of the intracellular leucine pool which, in turn, causes de-regulation of the enzymes encoded by the *leu* operon, resulting in ncBCAA biosynthesis.

The hypothesis was tested by comparing ncBCAA levels before and after IPTG-mediated induction of recombinant mini-proinsulin.

- *Hypothesis 7. Effective use of lac-based promoters in expression plasmids strongly depends on the levels of expressed LacI repressor and the E. coli strain employed as expression host.*

The P_{tac} promoter was used in this study for expression of recombinant mini-proinsulin from plasmid pSW3. The P_{tac} promoter is inducible by IPTG but, under certain conditions, promoter is leaky, i.e. expression is also triggered in a basal level even when no inducer is present. This mainly occurs when there are not sufficient LacI molecules to bind to the operators and repress basal expression.

The hypothesis was tested by comparing promoter efficiency of P_{tac} , i. e. levels of leaky expression and achieved recombinant protein production, cloned in pSW3 plasmid variants expressing different LacI repressor levels hosted in two different *E. coli* K-12 strains.

4. Materials & Methods

4.1 Materials, reagents and equipment

The different reagents, materials and equipment used in the current study are described in Table S1, Table S2 and Table S3, respectively.

4.2 Software

Software employed in this work is listed in Table S4.

4.3 Bacterial strains

Bacterial strains employed in this study are listed in Table S5.

4.4 Plasmids

An overview of plasmids used in this study and their genetic features is shown in Table S6. Plasmid maps were generated by Snapgene® software.

4.5 Media

Following media were used in this study:

- Luria-Bertani (LB) Medium (Roth, Cat.Nr: X964.1)
- Luria-Bertani (LB) Agar (Roth, Cat.Nr: X965.1)
- Super Optimal broth with Catabolite repression (SOC) (NEB, Cat.Nr.: B9020S)
- M9 medium

M9 medium contained (per 400 mL): 0.8 mL 1 M MgSO_4 , 0.4 mL 0.1 M CaCl_2 , 4 mL 40 % glucose, 8 mL 10 mg/mL thiamin, 16 mL 10 % casamino acids and 370 mL salt solution. The salt solution comprised (per L): 7.5 g $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 3 g KH_2PO_4 , 0.5 g NaCl and 1 g NH_4Cl .

- Modified Davis and Mingioli medium

Composition of modified Davis and Mingioli medium was as follows: 7 g/L K_2HPO_4 , 3 g/L KH_2PO_4 , 0.5 g/L Na citrate. $2\text{H}_2\text{O}$, 5 g/L $(\text{NH}_4)_2\text{SO}_4$, 4 g/L glucose, 0.1 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ and 100 $\mu\text{L/L}$ 10000 X trace elements solution. The 10000 X trace elements solution comprised (per 10 mL): 5 mL 0.1 M $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 0.2 mL 1 M $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.1 mL 1 M $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 0.1 mL 1 M $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.1 mL 0.2 M

CoCl₂·6H₂O, 0.2 mL 0.1 M CuCl₂·2H₂O, 0.1 mL 0.2 M NiCl₂·6H₂O, 0.2 mL 0.1 M Na₂MoO₄·5H₂O and 0.2 mL 0.1 M H₃BO₃.

- TUB medium

Composition of TUB mineral salt medium was as follows: 2 g/L Na₂SO₄, 2.468 g/L (NH₄)₂SO₄, 0.5 g/L NH₄Cl, 14.6 g/L K₂HPO₄, 3.6 g/L NaH₂PO₄·2H₂O and 1 g/L (NH₄)₂-H-citrate. The mineral salt medium was then supplemented with 2 mL/L MgSO₄ solution (1.0 M) and 2 mL/L trace elements solution. The trace element solution comprised (per L): 0.5 g CaCl₂ × 2H₂O, 0.18 g ZnSO₄ × 7H₂O, 0.1 g MnSO₄ × H₂O, 16.7 g FeCl₃ × 6H₂O, 0.16 g CuSO₄ × 5H₂O, 0.18 g CoCl₂ × 6H₂O.

4.6 Primers

Snapgene® software was employed in designing primers for both PCR and DNA sequencing purposes. Chemically synthesized primers were provided by ThermoFisher Scientific. An overview of the primers used in this study is summarized in Table S7.

4.7 Standard molecular biology methods

4.7.1 PCR

Two different DNA polymerases were employed, depending on the PCR objective. If PCR was destined to generate DNA fragments for subsequent cloning or sequencing reactions, the Phusion High-Fidelity DNA Polymerase was employed, since it has proofreading ability. PCR mix and thermocycling conditions when using Phusion High-Fidelity DNA Polymerase are shown in Table 9 and Table 10, respectively.

Table 9. PCR mix for Phusion® High-Fidelity DNA Polymerase.

PCR Component	50 µL reaction	Final concentration
Nuclease-free water	To 50 µL	-
5X Phusion HF or GC Buffer	10 µL	1X
10 mM dNTPs	1 µL	200 µM
10 µM Forward primer	2.5 µL	0.5 µM
10 µM Reverse primer	2.5 µL	0.5 µM
Template DNA	*Var.	-
DMSO (optional)	(1.5 µL)	3%
Phusion DNA Polymerase	0.5 µL	1 U/50 µL PCR
Total volume	50 µL	-

*Var.: Recommended amounts of DNA template for a 50 µL reaction are as follows:

DNA	Amount
Genomic	50 ng - 250 ng
Plasmid or viral	1 pg - 10 ng

Table 10. Thermocycling conditions for Phusion-based PCR. Ta: annealing temperature; ∞: infinite.

Step	Temperature (°C)	Time (min:s)	Number of cycles
Initial denaturation	98	0:30	1
Denaturation	98	0:10	5
Ta1	variable	0:30	
Extension	72	0:15 - 0:30 /kb	
Denaturation	98	0:10	30
Ta2	variable	0:30	
Extension	72	0:15 - 0:30 /kb	
Final extension	72	10:00	1
Hold	4	∞	

As opposed to standard primers, overlapping primers are primers containing a homologous region to template DNA as well as a non-homologous region at 5'. When using overlapping primers for PCR, two different annealing temperatures were employed. Ta1 was applied during the first 5 PCR cycles and was calculated taking into account only the homologous primer regions. Ta2 was applied in the last 30 PCR cycles and was calculated considering the whole primer sequence. If Ta1 and Ta2 were equal, same temperature was used for all cycles (routine PCR). Ext_t was calculated based on fragment size and nature of template. However, when PCR was performed for molecular clone verification purposes, where the exact sequence of the PCR product is not crucial, the so-called colony PCR variant was used in combination with the faster DreamTaq polymerase. In this case, a colony was resuspended in 20 µL of nuclease-free water and the mix was vortexed and heated at 95°C for 5 minutes to generate a cell lysate. Then, suspension was centrifuged for 2 min at 5000 g and 1 µL of supernatant was used as PCR template. PCR mix and thermocycling conditions when using DreamTaq DNA Polymerase are shown in Table 11 and Table 12, respectively. Nevertheless, some PCRs done in this study combined usage of both mentioned polymerases. In this case, thermocycling conditions of the DreamTaq polymerase were applied. PCRs were carried out with thermocycler SimpliAmp™ Thermal Cycler (Thermo Scientific). Calculation of annealing temperatures was performed by using software SnapGene (GSL Biotech LLC) and Tm Calculator (NEB).

Table 11. PCR mix for DreamTaq DNA Polymerase.

PCR component	50 µL reaction	Final concentration
Nuclease-free water	to 50 µL	-
10X DreamTaq Buffer	5 µL	1X
10 mM dNTPs	1 µL	200 µM

10 μ M Forward primer	2.5 μ L	0.5 μ M
10 μ M Reverse primer	2.5 μ L	0.5 μ M
Template DNA	*Var.	-
DreamTaq DNA Polymerase	0.625 μ L	1.25 U/50 μ L PCR
Total volume	50 μ L	-

*Var.: Recommended amounts of DNA template for a 50 μ L reaction are as follows:

DNA	Amount
Genomic	100 ng – 1 μ g
Plasmid or viral	10 pg – 1 ng

Table 12. Thermocycling conditions for colony PCR using DreamTaq polymerase. T_m: melting temperature; ∞ : infinite.

Step	Temperature ($^{\circ}$ C)	Time (min)	Number of cycles
Initial denaturation	95	3:00	1
Denaturation	95	0:30	40
Annealing	T _m – 5	0:30	
Extension	72	variable	
Final extension	72	15:00	1
Hold	4	∞	

4.7.2 Plasmid extraction (mini and midiprep)

A 5 mL LB medium containing the corresponding antibiotics was inoculated with a single colony of a certain *E. coli* strain. Culture was then incubated overnight at 30 or 37 $^{\circ}$ C and 220 rpm, overnight. Next day, plasmid purification was performed according to the protocol provided in Qiaprep spin miniprep kit (Qiagen, Ref: 27106). After plasmid isolation, DNA concentration and purity was measured by Nanodrop. However, for low copy plasmids, such as the pACG_araBAD and pACG_XylSPm plasmid variants, the QIAGEN Plasmid Plus Midi Kit (Qiagen; Ref.: ID12943) was employed. In that case, 50 mL LB containing the corresponding antibiotics was inoculated with a single colony of a certain *E. coli* strain. Culture was then incubated at 30 or 37 $^{\circ}$ C and 220 rpm, overnight. Next day, plasmid purification was performed by using the QIAGEN Plasmid Plus Midi Kit (Qiagen; Ref.: ID12943), according to the provided protocol. After plasmid isolation, DNA concentration and purity was determined by Nanodrop.

4.7.3 Genomic DNA extraction

Genomic DNA was extracted according to the following protocol, adapted from Packeiser *et al.* (2013). Cells from a colony were picked up from agar plates by using a sterile pipette tip and resuspended into 20 µl of 10 mM Tris / 1 mM EDTA (TE) buffer. To facilitate cell disruption, the mixture was vortexed for 10 s and incubated at 98 °C for 5 min. The lysate was next microcentrifuged (1 min, 10000 g) and the resulting supernatant was collected, diluted with distilled water at a 1:5~1:20 dilution ratio, and subjected to downstream applications.

4.7.4 Preparation of *E. coli* electro-competent cells

Depending on the exogenous DNA type to be transformed (linear dsDNA or plasmid DNA) and the features of the *E. coli* host (presence or not of plasmid pKD46), two different adapted protocols were employed in this study.

4.7.4.1 General protocol for plasmid electroporation into *E. coli* cells

20 mL of sterile LB medium were introduced into a 100 mL Erlenmeyer flask and medium was inoculated with a single colony of the corresponding *E. coli* host. Cells were incubated at 30-37 °C and 220 rpm, overnight. Next day, 100 mL of sterile LB medium were filled into a 500 mL Erlenmeyer flask and medium was inoculated with 1 mL of the fresh overnight culture. When the culture reached an OD_{600nm} of 0.4-0.5, cells were harvested as two 50 mL aliquots in sterile Falcon tubes and those were left on ice for 15 min. Tubes were then centrifuged for 15 min at 2500 g and 4 °C. After centrifugation, as much supernatant as possible was discarded and cell pellets were resuspended in 25 mL ice-cold dH₂O. Tubes were filled up to 50 mL with ice-cold dH₂O and cells were left on ice again for 15 min. Tubes were centrifuged again as described before. As much supernatant as possible was removed, pellet was resuspended in 25 mL ice-cold dH₂O, tubes were filled up to 50 mL and cells were incubated on ice for 15 minutes. Tubes were centrifuged again as described before. Now after discarding the supernatant, cells were resuspended in a total volume of 25 mL ice-cold dH₂O, both tubes were pooled into one tube, tube was filled up to 50 mL and cells were left on ice for 10 min. A last centrifugation step was performed as described before. Cells were resuspended in 1 mL ice-cold 10 % glycerol and 50 µL aliquots were prepared in sterile 1.5 mL Eppendorf tubes. Tubes were maintained in ice until electroporation took place or they were stored at -80 °C. Protocol was adapted from Reitz (2011).

4.7.4.2 Protocol for linear dsDNA transformation into *E. coli* cells containing plasmid pKD46

20 mL of sterile LB medium supplemented with 100 µg/mL ampicillin were introduced into a 100 mL Erlenmeyer flask and medium was inoculated with a single colony of *E. coli* BW25113 containing pKD46. Cells were incubated at 30 °C and 220 rpm, overnight. Next day, 100 mL of sterile LB medium supplemented with 100 µg/mL ampicillin were filled into a 500 mL Erlenmeyer flask and medium was inoculated with 1 mL of the fresh overnight culture. When the culture reached an OD_{600nm} of 0.1, expression of the phage red recombinase genes present in pKD46 were induced by adding 8 mL of a

1.5 M L-arabinose solution to reach a final arabinose concentration of 100 mM. Cells were then incubated again until the culture reached an OD_{600nm} of 0.7. Cells were harvested as two 50 mL aliquots in sterile Falcon tubes and those were left on ice for 15 min. Tubes were then centrifuged for 15 min at 2500 g and 4 °C. After centrifugation, as much supernatant as possible was discarded and cell pellets were resuspended in 25 mL ice-cold dH₂O. Tubes were filled up to 50 mL with ice-cold dH₂O and cells were left on ice again for 15 min. Tubes were centrifuged again as described before. As much supernatant as possible was removed, pellet was resuspended in 25 mL ice-cold dH₂O, tubes were filled up to 50 mL and cells were incubated on ice for 15 minutes. Tubes were centrifuged again as described before. Now after discarding the supernatant, cells were resuspended in a total volume of 25 mL ice-cold dH₂O, tubes were pooled into one tube, tube was filled up to 50 mL and cells were left on ice for 10 min. A last centrifugation step was performed as described before. Cells were resuspended in 1 mL ice-cold 10 % glycerol and 50 µL aliquots were prepared in sterile 1.5 mL Eppendorf tubes. Tubes were maintained in ice until electroporation took place or they were stored at -80 °C. Protocol was adapted from Reitz (2011).

4.7.5 Electroporation of DNA into *E. coli* electro-competent cells

A 1.5 mL Eppendorf tube containing 950 µL sterile SOC was prepared (for *E. coli* hosts containing pKD46, SOC medium was supplemented with 100 mM L-arabinose) and pre-warmed at 30-37 °C. A 2 mm electroporation cuvette was placed on ice and electroporator Gene Pulser Xcell™ (BioRad) was adjusted to the following conditions: 2500 V, 25 µF, 200 Ω. Electro-competent cells were thawed on ice (about 10 min) and 1-5 µL of the purified DNA were pipetted into an aliquot of competent cells. The resulting mix was homogenized by stirring with the pipette. The mix was transferred into an electroporation cuvette avoiding introduction of bubbles and making sure that cells deposit across the bottom of the cuvette. An electrical pulse was then applied. After transformation, the time constant was checked, which should be between 4.5 and 6.0. The pre-warmed SOC medium was pipetted into the cuvette immediately and the mix was homogenized gently with the pipette. The mix was then transferred to a 15 mL round-bottom tube. Cells were left incubating for 1 h at 37 °C (or 1.5-2 h at 30 °C) and 250 rpm, so that the antibiotic resistance gene can be expressed. 100 µL of the cell suspension (and/or serial dilutions) were streaked out on LB-plates containing the corresponding antibiotic for selection. Depending on the plasmid used for transformation, plates were incubated at 30 or 37 °C, overnight. Plates were stored at 4 °C. Protocol was adapted from Reitz (2011).

4.7.6 Curation of temperature-sensitive plasmids

Plasmids pKD46 and pCP20 are characterized by a temperature-sensitive origin of replication. In order to remove those plasmids from the cell, clones of interest containing one of those plasmids were streaked out for isolation of single colonies on a LB agar plate and it was incubated at 42 °C for 18 to 20 h. Then numerous single colonies were picked up from LB-plates and resuspended in 30 µL dH₂O. From the resulting suspension, a 0.5 µL droplet was pipetted in following selection plates: an LB plate containing ampicillin, an LB plate containing a second selective antibiotic, if required, and an LB plate without any antibiotic. Plates were incubated overnight at 37 °C. Full ampicillin sensitivity (no colony growth at all) indicated loss of plasmid pKD46 or pCP20. When no or partial ampicillin

sensitivity was shown (colony grew with difficulty), an additional temperature treatment at 42 °C was carried out, since plasmid rests could still remain in the cell.

4.7.7 Electrophoresis analytical gel

In this study, gel electrophoresis was used to verify amplification of target fragments after PCR and to check for appropriate restriction digestion. Therefore, 10 µL of sample (PCR or restriction products) were mixed with 2 µL Gel Loading Dye Purple (6X) (NEB, Ref: B7024). 10 µL of the resulting mix were then loaded on a E-Gel® 1.2 % Agarose (EtBr stained, Thermo Scientific, Cat.Nr.:G501801). 5 µL of the molecular marker Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb) (NEB, Ref: N0469) were used. DNA electrophoresis was run for 30 minutes at 100 V. Gel was then visualized using Gel Doc™ EZ Imager (BioRad).

4.7.8 PCR spin purification

QIAquick PCR Purification Kit (Qiagen; Ref.: 28104) was employed according to the provided protocol in order to purify PCR and restriction reaction products. This procedure performed better than gel purification since DNA loss during purification was much lower. However, this purification strategy was only employed when PCR product was really pure in the analytic gel, in order to avoid contamination with other non-desired DNA molecules.

4.7.9 Extraction from DNA gels

When the analytical gel revealed an unspecific PCR reaction (a smear of bands corresponding to PCR products of different sizes), purification of the target band corresponding to the correct PCR product was performed by extraction from a preparative gel. Hence, a 100 mL TAE buffer-based preparative gel containing 1% agarose and 0.5 µL 10 mg/mL ethidium bromide was prepared, DNA sample was loaded and gel was run at 180 V for 40 min. The target band was then cut and purified from the gel by using the QIAquick Gel Extraction Kit (Qiagen, Ref: 28706), according to the provided protocol.

4.7.10 Standard cloning based on restriction digestion and ligation

Restriction enzymes cleave DNA at or close to specific recognition sites. In this study, restriction enzymes were used for 3 different purposes: plasmid cloning, restriction analysis of potential clones and removal of PCR templates from PCR products by DpnI. DpnI is a restriction enzyme cutting methylated DNA. This enzyme is interesting when transforming linear PCR products into *E. coli*. DpnI treatment of PCR products ensures elimination of template plasmid DNA, thus reducing the likelihood of false positives on transformation plates. Restriction enzymes and its corresponding buffer were acquired from New England Biolabs (NEB). The setup for restriction digestion is shown in Table 13. The HF (high-fidelity)–variant of a restriction enzyme allows faster restriction (15 minutes at 37 °C) than the standard one (1 h at 37 °C).

Table 13. Setup of a standard restriction reaction. Incubation temperature, time and buffer is enzyme-dependent. RE: restriction enzyme.

Component	Amount for 50 μ L reaction
DNA	1 μ g
RE1-HF	0,5 μ L
RE2-HF	0,5 μ L
10X CutSmart Buffer	5 μ L
Nuclease-free water	Up to 50 μ L

After restriction of vector and insert with restriction enzymes for cloning purposes, T4 DNA Ligase was used to join both fragments. Ligation reaction setup was according to Table 14. Ligation was performed with the Quick Ligation Kit (NEB, Ref: M2200), according to the protocol provided. The standard 3:1 insert:vector molar ratio was employed in this study. Ligation mix was incubated at 25°C for 5 minutes, left on ice and 2 μ L of the ligation product were transformed into 50 μ L NEB 5-alpha competent *E.coli* (high efficiency) cells, according to the protocol provided (NEB, Ref: C2987H).

Table 14. Setup of a standard ligation reaction.

Component	Amount for 20 μ L reaction
Quick ligase reaction buffer 2X	10 μ L
Vector	100 ng
Insert	*
Nuclease-free water	Up to 20 μ L
Quick ligase	1 μ L

*Website 2 was used to calculate the optimal volume of insert necessary for ligation considering vector and insert sizes and the insert:vector ratio.

4.7.11 InFusion cloning

InFusion cloning is an advanced cloning approach allowing generation of complex DNA constructs by assembling multiple single fragments. The main advantage of this cloning strategy is that it does not require the usage of restriction enzymes or ligase. Single DNA fragments are joined and cloned into a linearized vector in a single and fast step. Each DNA fragment participating in the InFusion cloning reaction contains a 15 bp overlap homologous to the linearized vector or to the adjacent DNA fragment. Enzymes involved in the InFusion cloning reactions generate 15 bp ssDNA overhangs at the 5' end of both vector and DNA linear fragments which are then annealed at the complementary sites.

This novel cloning strategy was employed by using the In-Fusion® HD Cloning Kit (Takara; Ref.: 638909), according to the protocol provided. The InFusion cloning reaction was set up as shown in

Table 15. The InFusion cloning mix was incubated for 15 min at 50 °C and the resulting InFusion cloning product was transformed into competent *E. coli* NEB5 α cells (NEB; Ref.: C2987I), according to provided protocol.

Table 15. Setup of the InFusion cloning mix.

Compound	Volume (μ L)
5X InFusion HD Enzyme Premix	2
Linearized vector	*
Fragment 1	*
Fragment 2	*
Nuclease-free water	up to 10 μ L

*The on-line tool In-Fusion[®] Molar Ratio Calculator (Website 3) was used to calculate optimal amounts (volumes) of vector and insert for the In-Fusion[®] Cloning reaction considering fragment size and a relation vector:inserts of 1:2.

4.7.12 Mutagenesis

The kit “Phusion site-directed mutagenesis kit” (Thermoscientific, Ref: F-541) allows to introduce point mutations as well as deletions and insertions to available plasmids during PCR. In this strategy the entire original plasmid is amplified using 5'-phosphorylated primers which introduce the desired mutations at the target sites. The amplified linear PCR product, which now contains the desired mutation, is then autoligated in a 5-minute reaction.

The mix setup and thermocycling conditions employed in this study for the mutagenesis PCR reaction were as aforementioned in Table 9 and Table 10, respectively. Ligation was performed with the Quick Ligation Kit (NEB, Ref: M2200) by incubating ligation mix at 25 °C for 5 minutes and ligation product was transformed into 50 μ L NEB 5- α competent *E.coli* (high efficiency) cells according to the protocol provided (NEB, Ref: C2987H).

4.7.13 Sequencing

The target DNA fragment to be sequenced can be present in a plasmid or in a PCR product. When target DNA was contained in a plasmid, sequencing mix was prepared as indicated in Table 16.

Table 16. Setup of sequencing reaction using plasmid DNA as template.

Component	Volume (μ L)
Primer 10 μ M	1
Plasmid DNA (300 ng)	Variable
Nuclease-free water	Adjust to 6

When target DNA was present in a PCR product, 5 μ L of such PCR product were mixed with 2 μ L of Exo-Star Reagent (Sigma, Ref: GEUS78210) and the resulting mix was incubated for 15 minutes at 37 °C and then for 15 minutes at 80 °C. From the resulting mix 2 μ L were used for the sequencing reaction as described in Table 17. The ExoStar treatment, which comprises enzymes exonuclease I

and alkaline phosphatase, removes unbound primers and dNTPs from PCR products. Betaine improves sequencing performance of difficult DNA templates containing GC-rich base pairs, guanine stretches or GCT-type repeats (Haqqi *et al.*, 2002). Sequencing was performed internally at Sanofi-Aventis Deutschland GmbH.

Table 17. Setup of sequencing reaction using a PCR product as template.

Component	Volume (µL)
Primer 10 µM	1 µL
Exo-Star-treated PCR product	2 µL
Betaine 5M	2 µL
Nuclease-free water	1 µL

4.7.14 Preparation of bacterial glycerol stocks

A molecularly verified positive isolated colony of the corresponding strain was picked from a fresh LB plate and resuspended in a 4 mL LB liquid culture containing the corresponding antibiotic and it was grown at 30-37 °C at 220 rpm, overnight. 1 mL of sterile 50 % glycerol was added to 1 mL of the overnight culture in labelled 2 mL cryogenic vials. Cryovials were then stored at -80°C.

4.7.15 SDS-PAGE

Following components were mixed in order to prepare a 200 µL 1X sample buffer solution (per sample): 20 µL of 10X Bolt/NuPAGE Sample reducing agent, 50 µL Bolt/NuPAGE LDS sample buffer 4X and 130 µL dH₂O. OD1/mL cell pellets were resuspended with 200 µL 1X sample buffer to give a concentration of OD5/200 µL. Cell disruption was performed by sonication using VialTweeter block sonotrode attached to ultrasonic processor UP200St (standard settings: capacity: 100%, amplitude: 0.9 sec, duration: 1 min). Incubation was done for 10 minutes at 95°C. Afterwards, samples were shortly spinned and chilled down. 10-20 µL were loaded per lane (0.25-0.5 OD/lane) into a Bolt/NuPAGE Bis-Tris Puls gel containing 12% polyacrylamide. 10 µL of SeeBlue® Plus2 Pre-stained Protein Standard was also loaded. 1X Bolt/NuPAGE MES SDS Running Buffer was used as running buffer. 500 µL of Bolt/NuPAGE Antioxidant were loaded into the inner gel chamber. Electrophoresis was carried out at 190 V during 45 min. After electrophoresis, gel was left in a plastic tray and washed with dH₂O. Protein staining was then performed with Instant Blue (20 mL) for about 60 min, in agitation. Gel was again washed and de-stained for 3 times with dH₂O (each time 5-10 min, in agitation). Gel was then visualized using Gel Doc™ EZ Imager (BioRad).

4.8 Amino acid analysis by GC-FID

4.8.1 Intracellular soluble protein fraction and inclusion body isolation from cell extracts

The BugBuster Protein Extraction Reagent (Novagen, Ref: TB245) was used for isolation of Intracellular soluble protein and inclusion body fractions from cell extracts according to the following protocol:

1. OD_{600nm} of the culture was determined. A certain volume from each culture was taken and cells were harvested from liquid culture by centrifugation at 10000 g for 10 min. Supernatant was then decanted and removed from the pellet.
2. Cell pellet was resuspended in room temperature-BugBuster reagent by pipetting or gentle vortexing, using 5 mL reagent per gram of wet cell paste.
3. To reduce viscosity of the lysate and improve protein extraction, 10 µL Lysonase were added per gram wet cell paste. The cell suspension was mixed gently and it was incubated at room temperature on a shaking platform or rotating mixer at a slow setting for 20 minutes.
4. Insoluble cell debris was removed by centrifugation at 16000 g for 20 min at 4 °C. The supernatant was transferred to a fresh tube and it was maintained on ice for short term storage (2-3 h) or frozen at -20°C until needed. Supernatant corresponded to the intracellular soluble protein fraction, including free amino acids. Pellet was saved for further inclusion body purification.
5. Pellet was resuspended in the same volume of BugBuster Protein Extraction Reagent that was used to resuspend the original cell paste (step 2). Suspension was then pipetted up and down or vortexed. A complete resuspension of the cell pellet solubilizes and removes contamination proteins and is critical to obtain a high purity preparation.
6. 10 µL Lysonase were added per gram wet cell paste, suspension was mixed gently by vortexing and it was incubated at room temperature for 5 minutes.
7. An equal volume of 1:10 diluted BugBuster reagent (in deionized water) was added to the suspension and this was vortexed for 1 min.
8. The suspension was centrifuged at 5000g for 15 min at 4°C in order to collect the inclusion bodies. Supernatant was removed with a pipette.
9. The inclusion bodies were resuspended in 1:10 diluted BugBuster (10 mL per gram of original cell paste), vortexed and centrifuged as described in step 8.
10. Step 9 was repeated twice for a total of 3 washes with 1:10 diluted BugBuster.
11. Pellet was resuspended once more in 1:10 diluted BugBuster (10 mL per gram of original cell paste).
12. In order to check purification level of inclusion bodies containing the recombinant protein, an SDS-PAGE was carried out with the inclusion body samples. A certain volume corresponding to an OD1/mL was taken from the solution containing purified inclusion bodies obtained at step 11 and sample was mixed with 50 µL of Bolt LDS Sample Buffer (4X) and 20 µL of Bolt Reducing Agent (10X) and it was filled up to 200 µL with deionized water (OD5/200µL). Samples were heated at 70°C for 10 minutes and 20 µL were loaded into a Bolt Bis-Tris Puls gel containing 12% polyacrylamide gel (0.5 OD/lane). 10 µL of SeeBlue® Plus2 Pre-stained Protein Standard was also loaded. 1X Bolt MES SDS Running Buffer was employed as running buffer. 500 µL of Bolt Antioxidant were loaded into the inner gel chamber. Electrophoresis

was carried out at 190 V during 40 min. After electrophoresis, gel was left in a plastic tray and washed with dH₂O. Protein staining was then performed with Instant Blue (20 mL) for about 60 min, in agitation. Gel was again washed and de-stained for 3 times with dH₂O (each time 5-10 min, in agitation). Gel was then visualized using Gel Doc™ EZ Imager (BioRad).

13. The remaining IB suspension was centrifuged at 16000 g for 15 min at 4 °C and supernatant was discarded. Purified IB pellets were then stored at -80 °C.

4.8.2 Hydrolysis of intracellular soluble protein fraction and inclusion bodies

A given sample volume (max. 250 µL) was mixed up to 1 mL with 5M HCl. Resulting solutions were introduced in crystal vials with screw caps which are resistant to aggressive acids like HCl. Vials were incubated closed for 24h at 80 °C for acid hydrolysis. Afterwards, vials were left opened in a heating block for 16-24h at 65 °C while rotating until all liquid was evaporated. Hydrolyzed samples were then resuspended with dH₂O or a solution containing 20 mM HCl and 10% isopropanol and resulting solutions were used for further amino acid isolation.

4.8.3 ncBCAA analysis with EZ:faast™ for free (physiological) amino acid analysis by GC-FID

Amino acid isolation from samples was performed according to the following protocol, provided by the “EZ:faast™ for free (physiological) amino acid analysis by GC-FID” kit:

1. Preparation of Elution Medium: 3 volume parts of Reagent 3A were combined with 2 volume parts of Reagent 3B. For example, for 12 samples, 1.5 mL Reagent 3A and 1 mL Reagent 3B were mixed.
2. For calibration purposes, a certain volume of calibration standard solutions (SD1, SD2 and ncBCAA solutions) were mixed with 100 µL 200 µM L-2-aminobutyric acid (ABA). For analysis of samples, a certain volume of the hydrolyzed protein sample (inclusion body or intracellular protein soluble fraction) was mixed with 100 µL 200 µM L-2-aminobutyric acid (ABA). In order to identify retention times of canonical and non-canonical amino acids, a solution containing following components was additionally prepared for each GC-run set: 25 µL of 200 µM standard SD1, 25 µL of 200 µM standard SD2, 25 µL of 200 µM norvaline, 25 µL of 200 µM norleucine and 25 µL of 200 µM β-methylnorleucine. Standard solutions SD1 and SD2 were provided in the kit and contained canonical amino acids. ABA was used as internal standard amino acid in order to perform calculations.
3. Liquid samples generated at step 2 were introduced into provided sample preparation vials. A sorbent tip was then attached to a 1.5 mL syringe and piston was pulled back slowly (within 1 min) to pass total sample solution from the vial through sorbent tip.
4. 200 µL Reagent 2 (Washing Solution) were pipetted into the sample preparation vial.
5. Washing Solution was passed slowly through sorbent tip and into syringe barrel with the 1.5 mL syringe as well.
6. Sorbent tip with the bound amino acids was detached from the syringe and left it in the vial. Liquid accumulated in the syringe was discarded and syringe was washed with a 1:1 isopropanol:dH₂O solution for another use.

7. 200 μ L Elution Medium from step 1 were pipetted into the sample preparation vial.
8. A 0.6 mL syringe was taken and piston was pulled back halfway up the barrel. Sorbent tip with the bound amino acids was then attached to that syringe.
9. Elution Medium was passed slowly through the sorbent tip until liquid reached the white filter plug located in the tip.
10. Liquid and sorbent particles were then ejected out of the tip into the vial by pushing the piston down fast.
11. 50 μ L Reagent 4 (derivatization reagent) were pipetted with the provided Microdispenser into the vial.
12. The liquid was emulsified in the vial by vortexing for about 10-15 s and vial was left for 1 min at least in order to complete derivatization reaction. This step was repeated once.
13. 100 μ L Reagent 5 were pipetted into sample vial with the provided Microdispenser, solution was vortexed for 10-15 s and reaction was left to proceed for an additional minute.
14. 100 μ L Reagent 6 were pipetted into sample vial (without Microdispenser) and solution was vortexed for about 5 s. A centrifugation step at 3000 rpm for 1 min was performed to improve separation. At the end of the process, the emulsion should have separated into two layers. The upper layer contained the isolated amino acids.

4.8.4 ncBCAA analysis by gas chromatography-flame ionization detection (GC-FID)

After amino acid isolation process, around 120 μ L of the resulting upper layer were introduced into GC vials and 2 μ L were then injected into the GC analyzer. The GC was run according to following oven conditions: equilibration time of 0.5 min, 110 $^{\circ}$ C for 1 min, 30 $^{\circ}$ C/min heating up to 320 $^{\circ}$ C and then 320 $^{\circ}$ C for 1 min. Each GC run was about 9 min. Nitrogen was used as a carrier gas with a constant flow rate of 1.5 mL/min. Injection was carried out with a 1:15 split ratio at 250 $^{\circ}$ C.

4.8.5 Data analysis of data generated by GC-FID

For data analysis the software “Agilent Chemstation, Rev. B.03.02 [341]” was employed. For each GC run, a chromatogram containing peaks corresponding to different amino acids was generated. Most of the peaks were automatically detected and integrated by the software. However, some other smaller or more complex peaks, including the ones corresponding to the ncBCAA, had to be manually selected and integrated. After integration, a table was generated, where information and features about individual selected peaks (e.g. retention time, peak height, peak area,...) was summarized. ABA was selected as the internal standard for amino acid data analysis so that all individual amino acid calculations were normalized to ABA. Thus, peak area of each individual peak corresponding to one amino acid was divided by the peak area of ABA, obtaining a ratio. Those generated ratios were then interpolated into previously elaborated calibration curves and the equivalent amino acid concentration was estimated. From those concentrations, and knowing the volume of the samples, amino acid mass was then calculated. Those amino acid mass values could then be normalized to the corresponding OD_{600nm} of the cultivation as well as to the corresponding mass of expressed recombinant protein. In addition, ratio of normalized concentrations of canonical and the corresponding non-canonical BCAA could also be calculated.

4.9 Generation of pSW3 plasmid variants with different *lacI* expression levels: pSW3, pSW3_*lacI*⁺ and pSW3_*lacI*^q

A set of pSW3 plasmid variants were engineered, where different *lacI* promoter variants (*lacI*⁺ and *lacI*^q) were cloned together with gene *lacI* in order to express different amounts of LacI repressor (sections 4.9.1-4.9.3). The generated pSW3 plasmid variants were then transformed into *E.coli* K-12 BW25113 for further analysis of mini-proinsulin expression (4.9.4-4.9.6).

4.9.1 Sequencing of pSW3

The complete sequence of plasmid pSW3 was not available. However, after a large research through different patents and papers (Amann *et al.*, 1983; Habermann, 1986; Dörschung *et al.*, 1989; Habermann und Wengermayer, 1996), the sequence of some genetic regions present in plasmid pSW3 was identified. Using those sequences as templates, primers were accordingly generated in order to complete sequencing of the whole plasmid pSW3. Following primers were designed: pSW3_F1_seq, pSW3_R1_seq, pSW3_R2_seq, pSW3_F2_seq, pSW3_F3_seq, pSW3_R3_seq, pSW3_F4_seq, pSW3_R4_seq, pSW3_F5_seq, pSW3_R5_seq, pSW3_F6_seq, pSW3_R6_seq, pSW3_F7_seq, pSW3_R7_seq and pSW3_F8_seq. After sequencing, resulting sequencing reads were aligned (Figure S50) and the sequence and genetic map of whole pSW3 could be generated thanks to the multiple overlapping reads (Figure S2A). It was confirmed that pSW3 does not contain the *rop* gene. Hence, plasmid pSW3 is high-copy number (115) but not medium-copy number (39-55). Plasmid pSW3 lacks the *bom* site, which prevents transmission by conjugation, and contains the enhanced *lacI*^q promoter and an incomplete, non-functional *lacI* gene.

4.9.2 Cloning *lacI*^q into pSW3 to generate plasmid pSW3_*lacI*^q

The sequence of *lacI*⁺ and *lacI*^q promoter variants were obtained from Glascock and Weickert (1998) and the sequence of *lacI* was obtained from pGEX-4T-1 (Figure S4A). Cloning of *lacI*^q promoter variant and gene *lacI* into pSW3 was carried out through restriction with enzymes PaeI and PciI. First of all, fragment containing *lacI*^q promoter variant plus gene *lacI*, flanked by PaeI and PciI restriction sites, was synthetically generated by GeneArt, resulting in the plasmid 16ADFSUP_2034902_*lacI*^q_promotor (Figure S3A). Afterwards, restriction with PaeI and PciI of both pSW3 (vector) and 16ADFSUP_2034902_*lacI*^q_promotor (insert) was carried out and resulting fragments were ligated, thus resulting plasmid pSW3_*lacI*^q (Figure 8, Figure S2C). Ligation mix was then transformed into NEB 5-alpha competent *E.coli* (high efficiency) cells.

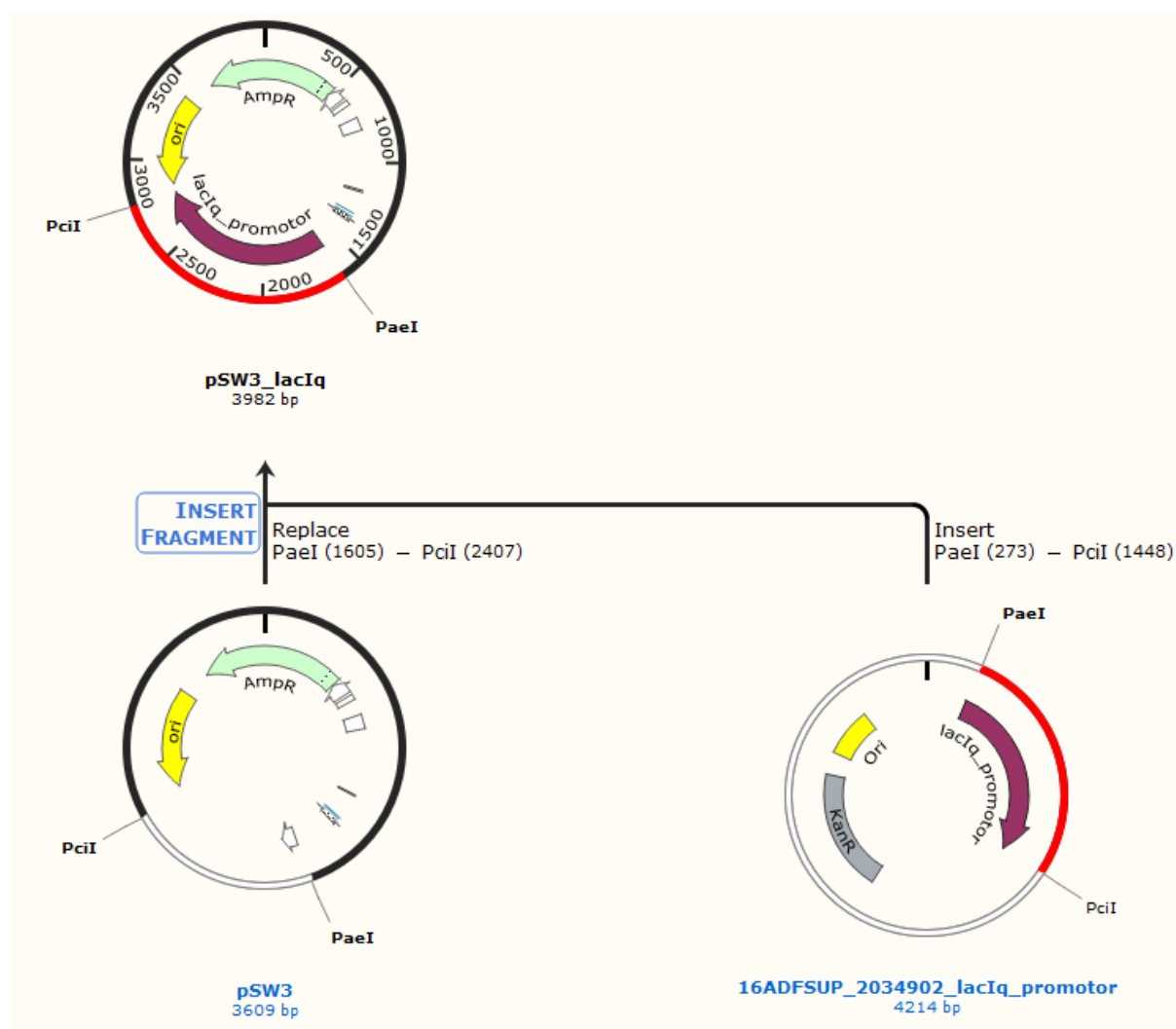


Figure 8. Cloning procedure followed to generate pSW3_lacI^q. Generated with Snapgene®.

20 colonies from the transformation plates were verified by colony PCR using primers pSW3_R3_seq and pSW3_R6_seq. PCR products were tested by gel electrophoresis (Figure S7), being all tested clones positive with the exception of clones 4, 9, 10, 12 and 13, which shown an additional non-expected band. In order to ensure that sequence of plasmids from positive clones was correct, some reactions were prepared for sequencing by using primers pSW3_R3_seq and pSW3_R6_seq and colony PCR product as template. pSW3_lacI^q sequence was correct for both tested clones 3 and 5 *E. coli* NEB5α pSW3_lacI^q (Figure S51). A bacterial glycerol stock was prepared for clone 3 *E. coli* NEB5α pSW3_lacI^q.

4.9.3 Mutation of pSW3_lacI^q to generate pSW3_lacI⁺

Plasmid pSW3_lacI⁺ (Figure S2B) was generated by introducing a point mutation into pSW3_lacI^q through PCR according to section 4.7.12. Following 5'-phosphorylated primers were used for mutagenesis PCR: Mut_lacI⁺_F and Mut_lacI⁺_R. Mutagenesis PCR was tested under different buffer combinations (buffer HF alone, buffer HF plus DMSO, buffer GC alone and buffer GC plus DMSO). Mutagenesis-PCR products were tested by gel electrophoresis (Figure S8). For the 4 tested PCR

conditions, the expected band was obtained. For the subsequent ligation step, PCR product obtained by using the HF buffer was employed. Ligation mix was then transformed into NEB 5-alpha competent *E.coli* (high efficiency) cells. 9 colonies from the transformation plates were verified by colony PCR using primers pSW3_R3_seq and pSW3_R6_seq. PCR products were tested by gel electrophoresis (Figure S9), being all tested clones positive. In order to certify if mutation in the *lacI* promoter region of the 9 tested clones occurred successfully, a reaction mix was prepared for sequencing. Primer pSW3_R3_seq was used as sequencing primer while colony PCR product was used as template. pSW3_*lacI*⁺ sequence was correct for clone 1 *E. coli* NEB5α pSW3_*lacI*⁺ (Figure S52). A bacterial glycerol stock was prepared for clone 1 *E. coli* NEB5α pSW3_*lacI*⁺.

4.9.4 Transformation of pSW3 into *E. coli* K-12 BW25113 and molecular verification

Electro-competent *E. coli* K-12 BW25113 cells were prepared and transformed with plasmid pSW3 by electroporation. In order to molecularly confirm transformation of plasmid pSW3, various clones growing in transformation plates were selected and a miniprep was done in order to purify plasmids. Verification of pSW3 presence was performed by restriction analysis using EcoRI. Digestion products were tested by gel electrophoresis (Figure S10), confirming that all 6 tested clones contained plasmid pSW3. A bacterial glycerol stock was prepared for clone 1 *E. coli* K-12 BW25113 pSW3.

4.9.5 Transformation of pSW3_*lacI*^q into *E. coli* K-12 BW25113 and molecular verification

Electro-competent *E. coli* K-12 BW25113 cells were prepared and transformed with plasmid pSW3_*lacI*^q by electroporation. Some colonies from the transformation plates were verified by colony PCR using primers pSW3_R3_seq and pSW3_R6_seq. PCR products were tested by gel electrophoresis (Figure S11), revealing that all 8 tested clones contained plasmid pSW3_*lacI*^q. A bacterial glycerol stock was prepared for clone 1 *E. coli* K-12 BW25113 pSW3_*lacI*^q.

4.9.6 Transformation of pSW3_*lacI*⁺ into *E. coli* K-12 BW25113 and molecular verification

Electro-competent *E. coli* K-12 BW25113 cells were prepared and transformed with plasmid pSW3_*lacI*⁺ by electroporation. Some colonies from the transformation plates were verified by colony PCR using primers pSW3_R3_seq and pSW3_R6_seq. PCR products were tested by gel electrophoresis (Figure S12), confirming that all 5 tested clones contained plasmid pSW3_*lacI*⁺. A bacterial glycerol stock was prepared for clone 1 *E. coli* K-12 BW25113 pSW3_*lacI*⁺.

4.10 Design of an arabinose-based tunable expression plasmid (pACG_araBAD)

A tunable expression plasmid under the control of the *araBAD* promoter system (pACG_araBAD) was engineered by Infusion cloning (sections 4.10.1-4.10.3) and genes *leuA*, *ilvC*, *ilvA*, *ilvIH*, *ilvBN*, *ilvGM* and *thrA* were subsequently cloned (section 4.10.4). Resulting plasmids would allow expression regulation of target genes by exogenous addition of L-arabinose.

4.10.1 Features of the DNA fragments involved in InFusion cloning

The plasmid pACG_araBAD was formed by the junction of 3 different DNA segments. Fragment 1 contained *araBAD* promoter, gene *araC*, multicloning site (MCS), His-tag sequence and T7-based terminator. Gene transcription was controlled by a tunable arabinose promoter (Guzman *et al.*, 1995) and a strong T7 terminator*. AraC is necessary for the activation of the arabinose promoter. The MCS allowed cloning with rare-cutting restriction enzymes NheI and NotI. Protein was expressed as a C-terminal 6xhis-tag fusion protein to allow detection by western blot. A gly-gly-gly-gly-ser linker was also available, connecting protein of interest with 6xhis-tag (Chen *et al.*, 2013). An efficient ribosome binding site from bacteriophage T7 gene 10 was also present in fragment 1 (Olins and Rangwala, 1989). Right after C-terminal 6xhis-tag three stop codons were present in order to ensure proper translation termination. Fragment 2 contained *cmR* cassette, which confers resistance against chloramphenicol. Fragment 3 contained origin of replication *ori2* and other genes involved in replication process. Those elements work coordinately, ensuring 1 copy number plasmid per cell, thus allowing cloning and stable maintenance of very large DNA fragments. Origin of replication was based on the Ori2 (OriS) replicon of the F (fertility) factor of *E. coli*, a vector which encodes the SopAB functions for active partitioning. These functions act at SopC to ensure that each daughter cell receives a copy of the plasmid. Initiation factor RepE (also known as RepA) mediates assembly of a replication complex at Ori2 (Imber *et al.*, 1983; Komori *et al.*, 1999; Mori *et al.*, 1986).

4.10.2 Generation of PCR fragments for InFusion cloning

The first step was to obtain by PCR the 3 abovementioned DNA segments either from plasmids already available or by chemical synthesis. Fragment 1 was chemically synthesized and subsequently cloned in plasmid 16ABZ5NP_1934177 (Figure S3B) by GeneArt. Original sequence of *araBAD* promoter as well as *araC* gene was obtained from plasmid pBAD-DEST49 (Figure S4C). Fragment 1 was then amplified by PCR from plasmid 16ABZ5NP_1934177 with primers araBAD_InFusion_F and araBAD_InFusion_R. Fragment 2 was directly amplified from plasmid pCP20 (Figure S1D) with primers cmR_InFusion_F and cmR_InFusion_R. Fragment 3 was directly amplified from pETcoco1 (Figure S4B) with primers ori2_InFusion_F and ori2_InFusion_R, respectively.

4.10.3 InFusion cloning to generate plasmid pACG_araBAD

Once PCR conditions were optimized for each DNA fragment, the three DNA segments were joint together according to the InFusion cloning strategy (Figure 9) to generate plasmid pACG_araBAD (Figure S5A).

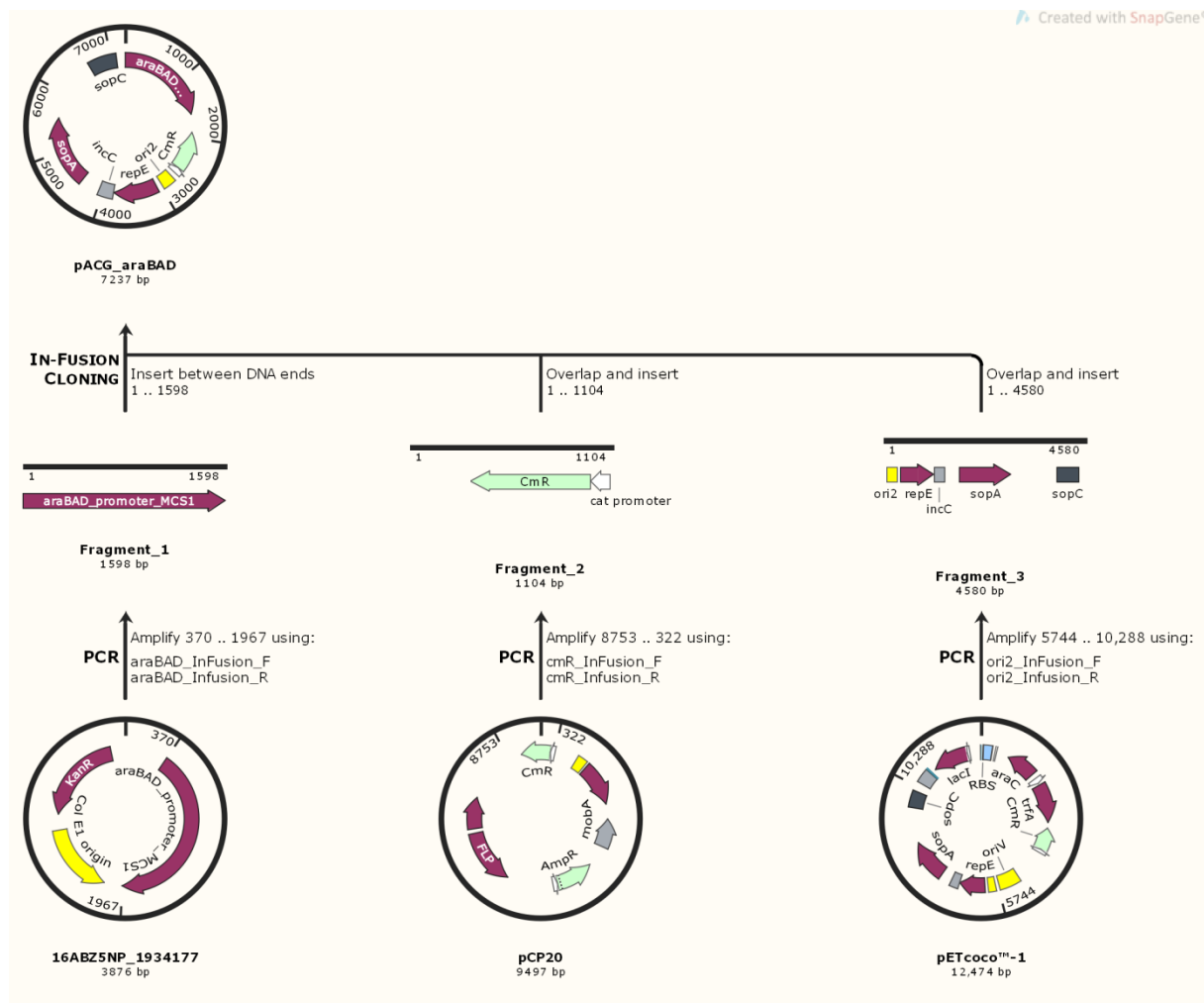


Figure 9. InFusion cloning procedure followed to generate pACG_araBAD plasmid. Fragment 1 was amplified by PCR from plasmid 16ABZ5NP_1934177 with primers araBAD_InFusion_F and araBAD_InFusion_R. Fragment 2 was directly amplified from plasmid pCP20 with primers cmR_InFusion_F and cmR_InFusion_R. Fragment 3 was directly amplified from pETcoco1 with primers ori2_InFusion_F and ori2_InFusion_R, respectively. The three DNA segments were then joint together according to the Infusion cloning strategy. Generated with Snapgene®.

The InFusion cloning product was transformed into NEB® 5-alpha Competent *E. coli* (High Efficiency) cells. Some clones growing on transformation plates were selected and a miniprep was done in order to purify plasmids. Verification of pACG_araBAD presence was performed by restriction analysis using enzymes XhoI and MssI. Digestion products were tested by gel electrophoresis (Figure S13), showing that all 7 tested clones contain plasmid pACG_araBAD. In order to ensure that sequence of plasmids from positive clones was correct, some reactions were prepared for sequencing by using primers InFusion1_seq_F1-F17 and InFusion1_seq_R1-R16 and plasmid preps as template. Sequence

of pACG_araBAD was correct for clone 1 *E. coli* NEB5 α pACG_araBAD (Figure S53). A bacterial glycerol stock was prepared for clone 1 *E. coli* NEB5 α pACG_araBAD.

4.10.4 Cloning of genes into plasmid pACG_araBAD

Once pACG_araBAD plasmid was verified, the following step was to amplify by PCR from the *E. coli* K-12 BW25113 genomic DNA the genes of study (*thrA*, *ilvA*, *leuA*, *ilvBN*, *ilvIH* and *ilvC*). Operon *ilvGM* was amplified from *E. coli* K-12 BW25113 *ilvG*⁺ (generated by Sarah Charaf) genomic DNA since *ilvG* is not functional in *E. coli* K-12 BW25113. Primers used for amplification of aforementioned genes contained NheI and NotI restriction sites at 5' and are summarized in Table 18. Subsequent cloning of the amplified genes into the tunable expression plasmid pACG_araBAD was carried out by restriction cloning using restriction enzymes NheI and NotI in order to yield plasmids pACG_araBAD_ geneX (Figure S5B-H). Plasmid pACG_araBAD_ *ilvBN* was generated and verified by Sarah Charaf.

Table 18. Primers and templates used for amplification of target genes from *E. coli* genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid pACG_araBAD.

Gene	Primers	Template
<i>thrA</i>	<i>thrA</i> _NheI_F and <i>thrA</i> _NotI_R	<i>E. coli</i> K-12 BW25113 genomic DNA
<i>ilvA</i>	<i>ilvA</i> _NheI_F and <i>ilvA</i> _NotI_R	<i>E. coli</i> K-12 BW25113 genomic DNA
<i>leuA</i>	<i>leuA</i> _NotI_F and <i>leuA</i> _NheI_R	<i>E. coli</i> K-12 BW25113 genomic DNA
<i>ilvGM</i>	<i>ilvGM</i> _NotI_F and <i>ilvGM</i> _NheI_R	<i>E. coli</i> K-12 BW25113 <i>ilvG</i> ⁺ genomic DNA
<i>ilvIH</i>	<i>ilvIH</i> _NheI_F and <i>ilvIH</i> _NotI_R	<i>E. coli</i> K-12 BW25113 genomic DNA
<i>ilvC</i>	<i>ilvC</i> _NheI_F and <i>ilvC</i> _NotI_R	<i>E. coli</i> K-12 BW25113 genomic DNA
<i>ilvBN</i>	<i>ilvBN</i> _NheI_F and <i>ilvBN</i> _NotI_R	<i>E. coli</i> K-12 BW25113 genomic DNA

Each cloning product was transformed into NEB® 5-alpha Competent *E. coli* (High Efficiency) cells. Some clones growing on transformation plates were selected and a miniprep was done in order to purify plasmids. Verification of plasmid presence was performed by restriction analysis as stated in Table 19. Digestion products were tested by gel electrophoresis (Figure S14-S18).

Table 19. Restriction enzymes used for verification of pACG_araBAD plasmid variants containing target genes (pACG_araBAD_ geneX) and the corresponding expected fragment sizes.

Plasmid	Restriction enzymes	Size of resulting fragments
pACG_araBAD	XhoI, MssI	2682 bp, 4555 bp
pACG_araBAD_ <i>thrA</i>	XhoI, NheI	3658 bp, 6022 bp
pACG_araBAD_ <i>ilvIH</i>	XhoI, NheI	3414 bp, 6022 bp
pACG_araBAD_ <i>ilvC</i>	XhoI, NheI	2671 bp, 6022 bp
pACG_araBAD_ <i>ilvA</i>	XhoI, NheI	2740 bp, 6022 bp

pACG_araBAD_ *leuA*

NotI, MssI

3040 bp, 5749 bp

Molecular verification of plasmid pACG_araBAD_ *ilvGM* was not performed by restriction analysis but by colony PCR. Some clones growing on transformation plates were selected and verified by colony PCR using primers InFusion1_seq_F4 and InFusion1_seq_R4. PCR products were tested by gel electrophoresis (Figure S19).

Afterwards, a midprep from clones verified by restriction analysis or PCR was carried out and some reactions were prepared for sequencing. Forward and reverse primers were designed so that they covered the plasmid region where the corresponding gene was cloned, hence generating overlapping reads (Table 20).

Table 20. Sequencing primers for verification of pACG_araBAD plasmid variants containing target genes (pACG_araBAD_ *geneX*).

pACG_araBAD_ <i>leuA</i>	pACG_araBAD_ <i>ilvC</i>	pACG_araBAD_ <i>ilvA</i>	pACG_araBAD_ <i>ilvIH</i>	pACG_araBAD_ <i>thrA</i>	pACG_araBAD_ <i>ilvGM</i>
			InFusion1_seq_ F4	InFusion1_seq_ F4	InFusion1_seq_ F4
			InFusion1_seq_ R4	InFusion1_seq_ R4	InFusion1_seq_ R4
InFusion1_seq_ F4	InFusion1_seq_ F4	InFusion1_seq_ F4	<i>ilvIH</i> _F1_seq	<i>thrA</i> _F1_seq	<i>ilvGM</i> _F1_seq
InFusion1_seq_ R4	InFusion1_seq_ R4	InFusion1_seq_ R4	<i>ilvIH</i> _F2_seq	<i>thrA</i> _F2_seq	<i>ilvGM</i> _F2_seq
<i>leuA</i> _F1_seq	<i>ilvC</i> _F1_seq	<i>ilvA</i> _F1_seq	<i>ilvIH</i> _F3_seq	<i>thrA</i> _F3_seq	<i>ilvGM</i> _F3_seq
<i>leuA</i> _F2_seq	<i>ilvC</i> _F2_seq	<i>ilvA</i> _F2_seq	<i>ilvIH</i> _F4_seq	<i>thrA</i> _F4_seq	<i>ilvGM</i> _F4_seq
<i>leuA</i> _F3_seq	<i>ilvC</i> _F3_seq	<i>ilvA</i> _F3_seq	<i>ilvIH</i> _F5_seq	<i>thrA</i> _F5_seq	<i>ilvGM</i> _F5_seq
<i>leuA</i> _R1_seq	<i>ilvC</i> _R1_seq	<i>ilvA</i> _R1_seq	<i>ilvIH</i> _R1_seq	<i>thrA</i> _R1_seq	<i>ilvGM</i> _R1_seq
<i>leuA</i> _R2_seq	<i>ilvC</i> _R2_seq	<i>ilvA</i> _R2_seq	<i>ilvIH</i> _R2_seq	<i>thrA</i> _R2_seq	<i>ilvGM</i> _R2_seq
<i>leuA</i> _R3_seq	<i>ilvC</i> _R3_seq	<i>ilvA</i> _R3_seq	<i>ilvIH</i> _R3_seq	<i>thrA</i> _R3_seq	<i>ilvGM</i> _R3_seq
			<i>ilvIH</i> _R4_seq	<i>thrA</i> _R4_seq	<i>ilvGM</i> _R4_seq
			<i>ilvIH</i> _R5_seq	<i>thrA</i> _R5_seq	

Plasmid sequence was correct for clone 2 *E. coli* NEB5α pACG_araBAD_ *leuA* (Figure S54), clone 4 *E. coli* NEB5α pACG_araBAD_ *ilvC* (Figure S55), clone 5 *E. coli* NEB5α pACG_araBAD_ *ilvA* (Figure S56), clone 3 *E. coli* NEB5α pACG_araBAD_ *thrA* (Figure S57), clone 1 *E. coli* NEB5α pACG_araBAD_ *ilvIH* (Figure S58) and clone 1 *E. coli* NEB5α pACG_araBAD_ *ilvGM* (Figure S59). Bacterial glycerol stocks were prepared for verified clones.

4.11 Design of an m-toluine-based tunable expression plasmid (pACG_XylSPm)

A tunable expression plasmid under the control of the XylS/Pm promoter system (pACG_XylSPm) was engineered, as an alternative to plasmid pACG_araBAD, due to its numerous advantages (section 4.11.1). Genes *leuA*, *ilvC*, *ilvA*, *ilvIH*, *ilvBN* and *thrA* were afterwards cloned (section 4.11.2). Resulting plasmids would allow expression regulation of target genes by exogenous addition of m-toluate.

4.11.1 Generation of plasmid pACG_XylSPm

Sequence of the XylS/Pm promoter region was obtained from pJB658 (Blatny *et al.*, 1997) (Figure S4D). This vector was proven to be successful for fine tuning expression in *E.coli* (Brautaset *et al.*, 1998; 2000; Winther-Larsen *et al.*, 2000; Sletta *et al.*, 2004). pJB658 was selected as a template due to its numerous advantages compared with other vectors of its type (Blatny *et al.*, 1997). In addition, pJB658 contains the unmodified *wild type* Pm promoter variant, which demonstrated to have a less basal expression than other mutagenized high-level expression variants (Balzer *et al.*, 2013; Binder *et al.*, 2016). The region of pJB658 containing XylS and Pm was chemically synthesized and subsequently cloned in plasmid 16ADCJKP_2028165_XylSPm (Figure S3C) by GeneArt, so that at both extremes of the cloned XylS/Pm promoter region MssI and NheI restriction sites were introduced. The XylS/Pm genetic region flanked by restriction sites was extracted from plasmid 16ADCJKP_2028165_XylSPm by cutting with NheI, PstI and MssI. The digestion product was loaded into a preparative gel and the band corresponding to the fragment containing NheI and MssI restriction sites at both ends (2174 bp) was cut and purified from the gel. In parallel, vector pACG_araBAD was also digested with NheI and MssI. Both vector and insert were ligated (Figure 10) to generate plasmid pACG_XylSPm (Figure S6A). The ligation product was transformed into NEB® 5-alpha Competent *E. coli* cells. Some clones growing on transformation plates were selected and molecular verification of plasmid pACG_XylSPm was performed by colony PCR using primers InFusion1_seq_F17 and InFusion1_seq_R4. PCR products were tested by gel electrophoresis (Figure S20), showing that clones 13, 16 and 19 were positive. Afterwards, a midiprep from PCR-verified clones was carried out and some reactions were prepared for sequencing by using primers InFusion1_seq_F17 and InFusion1_seq_R4, which covered the plasmid region where the XylS/Pm promoter was cloned. Sequence of plasmid pACG_XylSPm from clone 16 *E. coli* NEB5α pACG_XylSPm was successfully verified (Figure S60) and a bacterial glycerol stock was prepared.

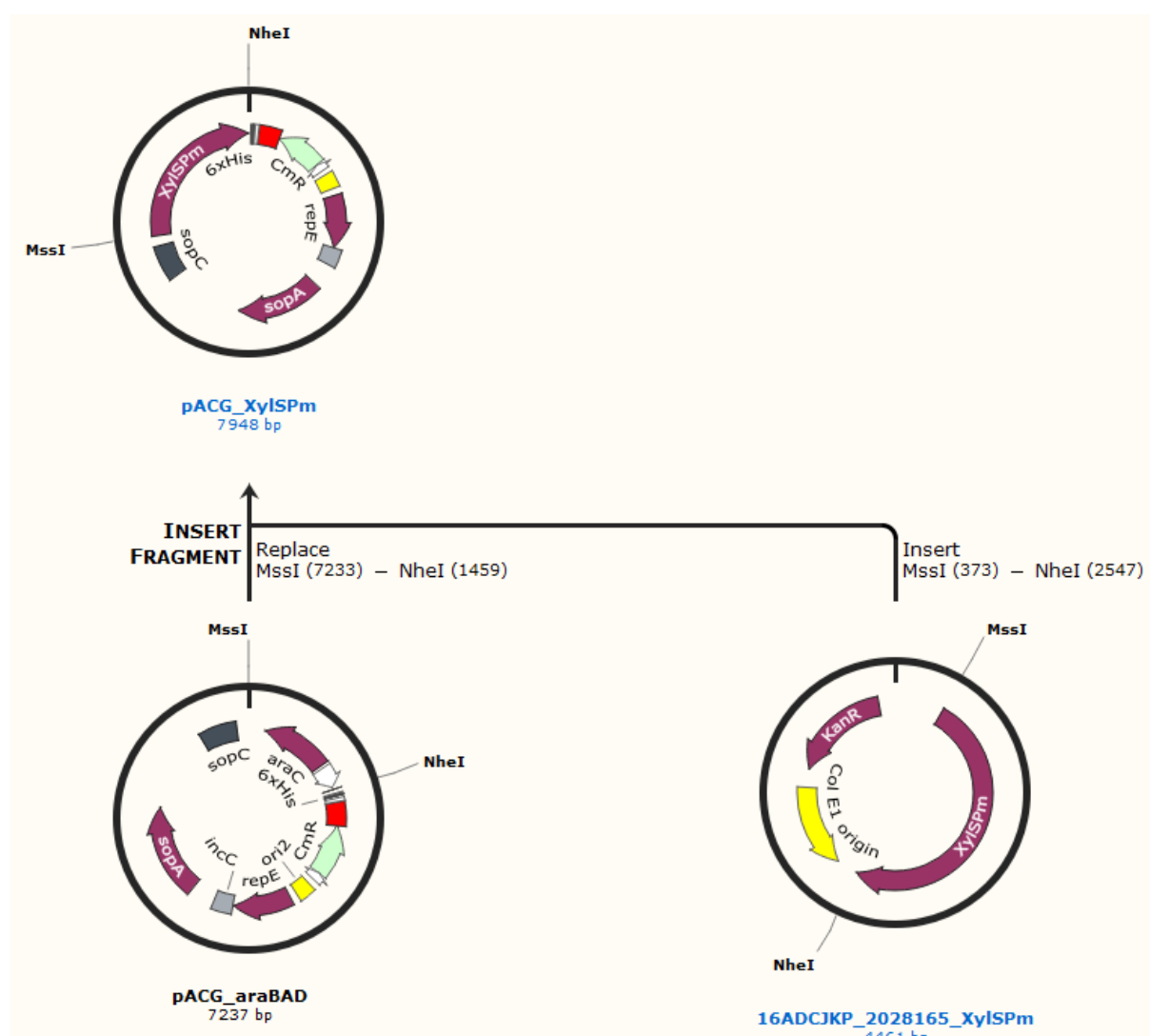


Figure 10. Cloning procedure followed to generate pACG_XylSPm plasmid. Generated with Snapgene®.

4.11.2 Cloning of genes into plasmid pACG_XylSPm

Once pACG_XylSPm plasmid was verified, the following step was to amplify by PCR from the *E. coli* K-12 BW25113 genomic DNA the genes of study (*thrA*, *ilvA*, *leuA*, *ilvIH*, *ilvBN* and *ilvC*). Primers used for amplification of aforementioned genes contained *NheI* and *NotI* restriction sites at 5' and are summarized in Table 18. Subsequent cloning of the amplified genes into the tunable expression plasmid pACG_XylSPm was carried out by restriction cloning using restriction enzymes *NheI* and *NotI* in order to yield plasmids pACG_XylSPm_geneX (Figure S6B-G). The cloning product was transformed into NEB® 5-alpha Competent *E. coli* cells. Molecular verification of the corresponding pACG_XylSPm_geneX plasmid was carried out by colony PCR. Some clones growing on transformation plates were selected and verified by colony PCR using primers InFusion1_seq_F17 and InFusion1_seq_R4. PCR products were tested by gel electrophoresis (Figure S21-S26). In order to ensure that sequence of plasmids from positive clones was correct, some reactions were prepared for sequencing by using colony PCR products as template and primers listed in Table 21.

Table 21. Sequencing primers for verification of pACG_XylSPm plasmid variants containing target genes (pACG_XylSPm_ geneX).

pACG_XylSPm	pACG_XylSPm	pACG_XylSPm	pACG_XylSPm	pACG_XylSPm	pACG_XylSPm
_leuA	_ilvC	_ilvA	_ilvIH	_thrA	_ilvBN
InFusion1	InFusion1	InFusion1	InFusion1	InFusion1	InFusion1
_seq_R4	_seq_R4	_seq_R4	_seq_R4	_seq_R4	_seq_R4
leuA_R1_seq	ilvC_R3_seq	ilvA_R3_seq	ilvIH_R5_seq	thrA_R5_seq	ilvBN_R2_seq
					ilvBN_F1_seq

Plasmid sequence was correct for clone 4 *E. coli* NEB5α pACG_XylSPm_ *leuA* (Figure S61), clones 1, 3, 6 and 9 *E. coli* NEB5α pACG_XylSPm_ *ilvC* (Figure S62), clone 21 *E. coli* NEB5α pACG_XylSPm_ *ilvIH* (Figure S63), clones 2 and 8 *E. coli* NEB5α pACG_XylSPm_ *thrA* (Figure S64), clone 3 *E. coli* NEB5α pACG_XylSPm_ *ilvBN* (Figure S65) and clones 4 and 6 *E. coli* NEB5α pACG_XylSPm_ *ilvA* (Figure S66) Bacterial glycerol stocks were prepared for verified clones.

4.12 Development of strains to regulate expression of the BCAA biosynthetic genes (*geneX*-tunable *E. coli*)

First of all, single knock-outs of the target endogenous genes involved in the BCAA synthesis were performed (sections 4.12.1 and 4.12.2). The respective electro-competent *E. coli* K-12 BW25113 Δ *geneX* mutants (*geneX*: *leuA*, *thrA*, *ilvA*, *ilvC*, *ilvIH*, *ilvBN* or *ilvGM*) were transformed with pSW3_ *lacI*⁺ (section 4.12.3), a high copy plasmid encoding mini-proinsulin, necessary for evaluation of ncBCAA mis-incorporation; and a tunable expression plasmid variant pACG_araBAD_ *geneX* (section 4.12.4) or pACG_XylSPm_ *geneX* (section 4.12.5), a 1-copy plasmid allowing expression regulation of target genes, previously knocked-out (Figure 11).

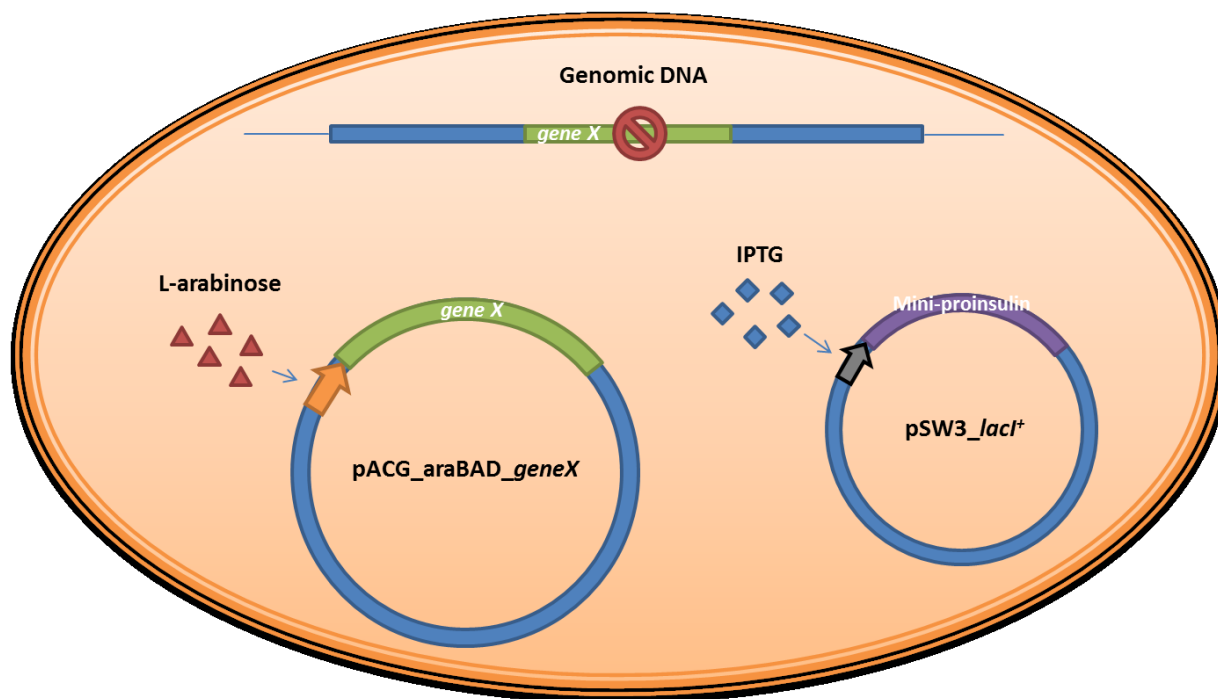


Figure 11. Overview of the genetic modifications carried out in this study in order to generate *geneX*-tunable *E. coli* strains. Each engineered *geneX*-tunable *E. coli* lacked a certain endogenous target gene, generally indicated as *geneX* in this work (*E. coli* K-12 BW25113 $\Delta geneX$), and contained one plasmid allowing expression regulation of such target gene by exogenous L-arabinose (pACG_araBAD_ *geneX*) and one plasmid enabling IPTG-mediated expression of recombinant mini-proinsulin for subsequent analysis of the impurity profile (pSW3_ *lacI*⁺). A total of seven tunable *E. coli* strains were engineered, one for each target gene: *leuA*, *thrA*, *ilvA*, *ilvC*, *ilvIH*, *ilvBN* or *ilvGM*.

4.12.1 Generation of strains *E. coli* K-12 BW25113 $\Delta geneX$ (*geneX*: *leuA*, *thrA*, *ilvA* and *ilvC*)

Strain *E. coli* K-12 BW25113 as well as single knock-out mutants *E. coli* K-12 BW25113 *leuA:kanR*, *E. coli* K-12 BW25113 *thrA:kanR*, *E. coli* K-12 BW25113 *ilvA:kanR* and *E. coli* K-12 BW25113 *ilvC:kanR* containing pKD46 were acquired from the *E. coli* Genetic Stock Center (CGSC) from Yale University. Those mutant strains belong to the so-called KEIO collection (Baba *et al.*, 2006). The genome of those mutant strains contains a kanamycin resistance marker substituting the target gene. CGSC identification for each acquired strain is indicated in Table S5.

Plasmid pKD46 was curated from the acquired *E. coli* K-12 BW25113 single knock-out mutants (section 4.12.1.1) and the respective electro-competent cells were transformed with pCP20, a temperature-sensitive plasmid encoding a flipase, in order to trigger removal of the kanamycin cassette from the genome by FRT-specific recombination (sections 4.12.1.2 and 4.12.1.3). Plasmid pCP20 was then curated and removal of the antibiotic resistance marker from *E. coli* K-12 BW25113 single mutants was tested by colony PCR (section 4.12.1.4).

4.12.1.1 Curation of pKD46

Curation of plasmid pKD46 was performed according to section 4.7.6 (plates not shown). In order to molecularly confirm removal of plasmid pKD46, one of the clones growing in LB and LB + kanamycin plates was selected for each *E. coli* K-12 BW25113 *geneX:kanR* tested strain and a miniprep was done in order to purify plasmids. Verification of pKD46 presence was performed by restriction analysis using EcoRI. Digestion products were tested by gel electrophoresis (Figure S27), showing that all tested clones lacked pKD46.

4.12.1.2 Transformation of *E. coli* K-12 BW25113 *geneX:kanR* with pCP20 and removal of the kanamycin resistant cassette

A miniprep of an *E. coli* BT340 (CGSC#7629) culture was done in order to isolate plasmid pCP20. Electro-competent *E. coli* K-12 BW25113 *geneX:kanR* cells were prepared and transformed with plasmid pCP20 by electroporation. In order to molecularly confirm transformation of plasmid pCP20, 2 clones growing in transformation plates were selected for each *E. coli* K-12 BW25113 *geneX:kanR* tested strain (for *E. coli* K-12 BW25113 *ilvI:kanR* 4 clones were selected) and a miniprep was done in order to purify plasmids. Verification of pCP20 presence was performed by restriction analysis using EcoRI. Digestion products were tested by gel electrophoresis (Figure S28), showing that all tested clones contained pCP20.

4.12.1.3 Verification of kanamycin-resistance cassette removal from genomic DNA of *E. coli* K-12 BW25113 $\Delta geneX$ mutants

In order to molecularly confirm the flippase-mediated removal of kanamycin-resistance cassette from genomic DNA in *E. coli* K-12 BW25113 $\Delta geneX$ mutants containing pCP20, 2 clones were selected for each *E. coli* K-12 BW25113 $\Delta geneX$ strain (for *E. coli* K-12 BW25113 $\Delta ilvI$ 3 clones were selected) and were subjected to colony PCR using primers listed in Table 22. PCR products were tested by gel electrophoresis (Figure S29), showing that for all tested clones kanamycin resistance marker was successfully removed from the genome.

Table 22. Primers used for PCR-verification of flippase-mediated removal of the kanamycin resistance cassette from the genome of strains *E. coli* K-12 BW25113 $\Delta geneX$ pCP20 and the corresponding expected PCR product sizes for both kanS and kanR variants.

Strain to verify	Primers	Expected size PCR product (bp)
<i>E. coli</i> K-12 BW25113 $\Delta thrB$ pCP20	thrB_F, thrB_R	kanS ¹ : 1081, kanR ² : 2281
<i>E. coli</i> K-12 BW25113 $\Delta thrC$ pCP20	thrC_F, thrC_R	kanS: 691, kanR: 1891
<i>E. coli</i> K-12 BW25113 $\Delta ilvA$ pCP20	ilvA_F, ilvA_R	kanS: 727, kanR: 1927
<i>E. coli</i> K-12 BW25113 $\Delta ilvB$ pCP20	ilvB_F, ilvB_R	kanS: 1307, kanR: 2507
<i>E. coli</i> K-12 BW25113 $\Delta thrA$ pCP20	thrA_F, thrA_R	kanS: 1164, kanR: 2364

<i>E. coli</i> K-12 BW25113 $\Delta leuA$ pCP20	<i>leuA_F</i> , <i>leuA_R</i>	kanS: 746, kanR: 1946
<i>E. coli</i> K-12 BW25113 $\Delta ilvC$ pCP20	<i>ilvC_F</i> , <i>ilvC_R</i>	kanS: 706, kanR: 1906
<i>E. coli</i> K-12 BW25113 $\Delta ilvI$ pCP20	<i>ilvIH_F</i> , <i>ilvIH_R</i>	kanS: 1246, kanR: 2446, wild type: 2844

1: kan-sensitive clone (kanS), i.e. clone where kan resistance marker was removed from the genome.

2: kan-resistant clone (kanR), i.e. clone where kan resistance marker still remains integrated into the genome.

4.12.1.4 Curation of pCP20

Curation of plasmid pCP20 was performed according to section 4.7.6. For strains *E. coli* K-12 BW25113 $\Delta leuA$, $\Delta ilvC$ and $\Delta ilvI$ a second temperature treatment of 42 °C was necessary to completely remove plasmid pCP20. Plate results are shown in Figure S67. A bacterial glycerol stock was prepared for clone 2.3.1 *E. coli* K-12 BW25113 $\Delta leuA$, clone 2.2.1 *E. coli* K-12 BW25113 $\Delta ilvC$, clone 1.2.1 *E. coli* K-12 BW25113 $\Delta ilvI$, clone 1.1 *E. coli* K-12 BW25113 $\Delta ilvA$, clone 1.1 *E. coli* K-12 BW25113 $\Delta ilvB$, clone 1.1 *E. coli* K-12 BW25113 $\Delta thrA$, clone 1.1 *E. coli* K-12 BW25113 $\Delta thrB$ and clone 1.1 *E. coli* K-12 BW25113 $\Delta thrC$.

4.12.2 Generation of strains *E. coli* K-12 BW25113 $\Delta ilvIH$ and *E. coli* K-12 BW25113 $\Delta ilvBN$

The knock-out strains *E. coli* K-12 BW25113 $\Delta ilvIH$ and $\Delta ilvBN$ were not acquired from CGSC but generated in the laboratory. Strain *E. coli* K-12 BW25113 $\Delta ilvIH$ was generated in the context of this thesis while strain *E. coli* K-12 BW25113 $\Delta ilvBN$ was generated by Sarah Charaf. A summary of the procedure carried out to generate strain *E. coli* K-12 BW25113 $\Delta ilvIH$ is described below. Electro-competent *E. coli* K-12 BW25113 cells were transformed with pKD46, a temperature-sensitive helper plasmid, expressing the elements enabling homologous recombination (section 4.12.2.1). Knock-out mutants for the operons *ilvIH* and *ilvBN* were then generated by transformation of electro-competent *E. coli* K-12 BW25113 cells containing pKD46 with the respective deletion cassette, previously obtained by PCR from pKD3 or pKD4. PCR-based verification was carried out to test proper integration of the deletion cassette into the genome (section 4.12.2.2). Plasmid pKD46 was curated (section 4.12.2.3) and the respective electro-competent *E. coli* K-12 BW25113 mutants were transformed with pCP20, a temperature-sensitive plasmid encoding a flipase, in order to trigger removal of the antibiotic cassette from the genome by FRT-specific recombination (section 4.12.2.4). Plasmid pCP20 was curated and removal of the antibiotic resistance marker from *E. coli* K-12 BW25113 $\Delta ilvIH$ and $\Delta ilvBN$ single mutants was tested by sequencing (section 4.12.2.5).

4.12.2.1 Transformation of *E. coli* K-12 BW25113 with pKD46

Electro-competent *E. coli* K-12 BW25113 cells were prepared and transformed with plasmid pKD46 by electroporation. Six clones growing on transformation plates were selected and a miniprep was done in order to purify plasmids. Verification of pKD46 presence was performed by restriction analysis using *EcoRI*. Digestion products were tested by gel electrophoresis (Figure S30), showing that all 6 tested clones contained pKD46.

4.12.2.2 Transformation of *E. coli* K-12 BW25113 pKD46 with the *ilvIH*-deletion cassette

The *ilvIH* knock-out cassette was generated by PCR amplification of the chloramphenicol or kanamycin resistance cassette with primers KO3_F and KO3_R, by using as template plasmid pKD3 or pKD4, respectively. PCR primers were designed so that they contained 36 bp homologous to the genomic region flanking operon *ilvIH* at 5' and 20 bp homologous to priming sites 1 and 2 of the template plasmids at 3' (Table 23). PCR was performed with Phusion DNA Polymerase by using different buffer combinations (buffer HF or GC plus DMSO) and PCR products were tested by gel electrophoresis (Figure S31).

Table 23. Primers used to generate the *ilvIH* knock-out cassette. Highlighted in yellow at 5' are 36 bp homologous to the genomic region flanking operon *ilvIH*. At 3', 20 bp homologous to priming sites 1 and 2 present in plasmids pKD3 and pKD4 are highlighted in blue and green respectively.

Primer name	Primer sequence (5' – 3')
KO3_F	TTACACATTTTTCGTCAAACAGTGAGGCAGGCCGTGTAGGCTGGAGCTGCTTC
KO3_R	ACATGTTGGGCTGTAAATTGCGCATTGAGATCATTCATGGGAATTAGCCATGGTCC

A preparative gel was then prepared and 200 µL of the PCR product generated by using buffer GC plus DMSO and pKD3 as template were loaded. After electrophoresis, the corresponding band of 1105 bp was cut from the gel, DNA was purified and concentration was determined by Nanodrop. Purified PCR product was additionally treated with DpnI in order to remove any rest of plasmid DNA template. Electro-competent *E. coli* K-12 BW25113 pKD46 cells were prepared and transformed with the purified *ilvIH* knock-out cassette by electroporation. In order to molecularly confirm *ilvIH* knock out, genomic DNA of 10 colonies grown on the transformation plates was extracted and PCR was carried out with primers *ilvIH*_F and *ilvIH*_R. PCR products were tested by gel electrophoresis (Figure S32), being clone 1 the only positive clone.

4.12.2.3 Curation of pKD46

Curation of plasmid pKD46 from clone 1 *E. coli* K-12 BW25113 *ilvIH:cmR* pKD46 was performed according to section 4.7.6. In this case, a second temperature treatment of 42 °C was necessary to completely remove plasmid pKD46. Plate results are shown in Figure S68 and indicated that all potential subclones originating from clone 1 *E. coli* K-12 BW25113 *ilvIH:cmR* pKD46 lost plasmid pKD46 after temperature treatment.

4.12.2.4 Transformation of *E. coli* K-12 BW25113 *ilvIH:cmR* with pCP20 and removal of the chloramphenicol resistant cassette

Electro-competent *E. coli* K-12 BW25113 *ilvIH:cmR* cells were prepared and transformed with plasmid pCP20 by electroporation. In order to molecularly confirm if flippase-mediated site specific recombination took place and chloramphenicol resistant marker was removed from the genome, genomic DNA of 4 colonies grown on the transformation plates was extracted and PCR was carried

out with primers *ilvIH_F2* and *ilvIH_R*. PCR products were tested by gel electrophoresis (Figure S33), showing that all evaluated clones were positive.

4.12.2.5 Curation of pCP20

Curation of plasmid pCP20 was performed according to section 4.7.6. In this case, a second temperature treatment of 42 °C was necessary to completely remove plasmid pCP20. Plate results are shown in Figure S69, concluding that all potential clones lost plasmid pCP20.

In order to ensure that *E. coli* K-12 BW25113 $\Delta ilvIH$ mutant was correct, sequencing of the mutated region was performed. Hence, a PCR of the genomic region flanking the *ilvIH* deletion was carried out with primers *ilvIH_F2* and *ilvIH_R* and the resulting PCR product was sent for sequencing. Sequencing reactions were done with the same primers as for PCR. According to sequencing results, the sequence on the *ilvIH*-deleted region in *E. coli* K-12 BW25113 $\Delta ilvIH$ was:
tttctttcacctttcctcctgtttattcttattaccccggtttatgtctctggctgccaattgcttaagcaagatcggacggtaatgtgtttacacat
ttttccgtcaaacagtgaggcaggccgtgtaggctggagctgcttgaagttcctatactttctagagaataggaacttcggaataggaactaa
ggaggatattcatatggaccatggctaattcccatgaatgatctcaatgcgcaatttacagcccaacatgtcacgttgggcttttttgcgaaatca
gtgggaacctggaataaaagcagttgccgcagtttaattttctgcgcttagatgttaatgaatt. Highlighted in green appears the sequence located upstream from the *ilvIH* deletion while in red appears the sequence located downstream. The scar segment is displayed in yellow, with the FRT site in bold. After sequence verification, a bacterial glycerol stock of strain *E. coli* K-12 BW25113 $\Delta ilvIH$ was prepared.

4.12.3 Transformation of pSW3_ *lacI*⁺ into *E. coli* K-12 BW25113 $\Delta geneX$

Electro-competent *E. coli* K-12 BW25113 $\Delta geneX$ cells were prepared and transformed with the previously verified plasmid pSW3_ *lacI*⁺ by electroporation. In order to molecularly confirm presence of plasmid pSW3_ *lacI*⁺, some clones growing on transformation plates were selected for each *E. coli* K-12 BW25113 $\Delta geneX$ tested strain and were subjected to colony PCR using primers pSW3_R3_seq and pSW3_R6_seq. PCR products were tested by gel electrophoresis (Figure S34-S37). According to the gels all clones were positive, being confirmed that they contained plasmid pSW3_ *lacI*⁺. A bacterial glycerol stock was prepared with a verified clone from each tested strain.

4.12.4 Transformation of pACG_araBAD_ *geneX* into *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ (*geneX*-tunable *E. coli*)

Electro-competent *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ cells were prepared and transformed with the previously verified plasmid pACG_araBAD_ *geneX* by electroporation. In order to molecularly confirm presence of plasmid pACG_araBAD_ *geneX*, some clones growing on transformation plates were selected for each *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ tested strain and were subjected to colony PCR using primers InFusion1_seq_F4 and InFusion1_seq_R4. PCR products were tested by gel electrophoresis (Figure S38-S44). According to the gels all clones were positive, being confirmed that they contained plasmid pACG_araBAD_ *geneX*. A bacterial glycerol stock was prepared with a verified clone from each tested strain. The generated strains would allow transcription regulation of single

BCAA biosynthetic genes by exogenous L-arabinose addition. Through this dissertation the term *geneX-tunable E. coli* is often used instead of the complete strain name (*E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ pACG_araBAD_ *geneX*) for simplification reasons.

4.12.5 Transformation of pACG_XylSPm_ *geneX* into *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺

Electro-competent *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ cells were prepared and transformed with the previously verified plasmid pACG_XylSPm_ *geneX* by electroporation. In order to molecularly confirm presence of plasmid pACG_XylSPm_ *geneX*, some clones growing on transformation plates were selected for each *E. coli* K-12 BW25113 $\Delta geneX$ tested strain and were subjected to colony PCR using primers InFusion1_seq_F17 and InFusion1_seq_R4. PCR products were tested by gel electrophoresis (Figure S45-S49). According to the gels all clones were positive, being confirmed that they contained plasmid pACG_XylSPm_ *geneX*. A bacterial glycerol stock was prepared with a verified clone from each tested strain. The generated strains would allow transcription regulation of single BCAA biosynthetic genes by exogenous m-toluate addition.

5. Results

5.1 Analysis of mini-proinsulin expression in *E. coli* K-12 W3110M and *E. coli* K-12 BW25113 containing plasmid pSW3

Together with the novel Pm/XylS promoter, the classic *araBAD* promoter was also utilized in this thesis in order to regulate expression of target genes involved in the BCAA biosynthetic pathway. In order to ensure functionality of the *araBAD* promoter, the host strain must be deficient in arabinose catabolizing enzymes. For this reason, strain *E. coli* K-12 BW25113 was selected as the model organism for this study, since it contains a deletion of the endogenous *araBAD* operon. However, the plasmid expressing recombinant mini-proinsulin under the control of a P_{tac} promoter selected for this study (pSW3) was just previously tested on strain *E. coli* K-12 W3110M, showing here good levels of expression and low promoter leakiness. Hence, it was necessary to evaluate first performance of plasmid pSW3 in the background of *E. coli* K-12 BW25113.

The aim of this experiment was to evaluate expression of recombinant mini-proinsulin in both strains *E. coli* K-12 W3110M and *E. coli* K-12 BW25113 containing plasmid pSW3 by SDS-PAGE. It was aimed to check how leaky was the expression of the P_{tac} promoter by using two different *E. coli* K-12 genetic backgrounds (W3110M and BW25113) and to confirm if, unlike *E. coli* K-12 W3110M (*lacI^s*), *E. coli* K-12 BW25113 is a *lacI⁺* strain.

5.1.1 Cultivation conditions

Pre-culture was prepared as follows: 10 mL of M9 medium were introduced in sterile 100 mL Erlenmeyer shake flasks. One flask was prepared for each tested strain (*E. coli* K-12 W3110M, *E. coli* K-12 W3110M pSW3, *E. coli* K-12 BW25113 and *E. coli* K-12 BW25113 pSW3). Ampicillin was supplemented to a final concentration of 100 µg/mL for strains containing pSW3. Fresh cells were inoculated into the medium and cultures were incubated at 37 °C und 220 rpm, overnight. The main culture was then prepared and cultivated as follows: 8 sterile 300 mL Erlenmeyer shake flasks were filled with 30 mL M9 medium. 2 shake flasks were prepared for each tested strain. OD_{600nm} of pre-cultures was measured and the volume from the pre-culture needed to generate a main culture at 0.05 OD_{600nm} in a final volume of 30 mL was calculated. The corresponding calculated volume of pre-culture needed was introduced in the 300 mL sterile Erlenmeyer shake flasks. Ampicillin was supplemented to a final concentration of 100 µg/mL for strains containing pSW3. Resulting cultures were incubated at 37°C and 220 rpm until OD_{600nm} reached 0.5-0.8. Then, 300 µL of 40% glucose were added to all cultures while IPTG was added to a concentration of 0.5 mM only to one of the shake flasks available for each tested strain (induced sample). The other shake flask was used as negative control for induction (non-induced sample). Cultures were then left incubating at 37°C and 220 rpm. Around 20h after induction, OD_{600nm} was measured and the corresponding volume needed to generate a 1 mL solution at OD_{600nm} =1 was taken out, centrifuged (6000g, 5min) and the resulting pellet was stored at -20°C for SDS-PAGE analysis.

5.1.2 Mini-proinsulin analysis with SDS-PAGE

Mini-proinsulin expression was evaluated by SDS-PAGE with samples taken 20 h after IPTG induction. For each tested strain, 2 cultivations were evaluated: one with IPTG induction and the other, without induction (Figure 12).

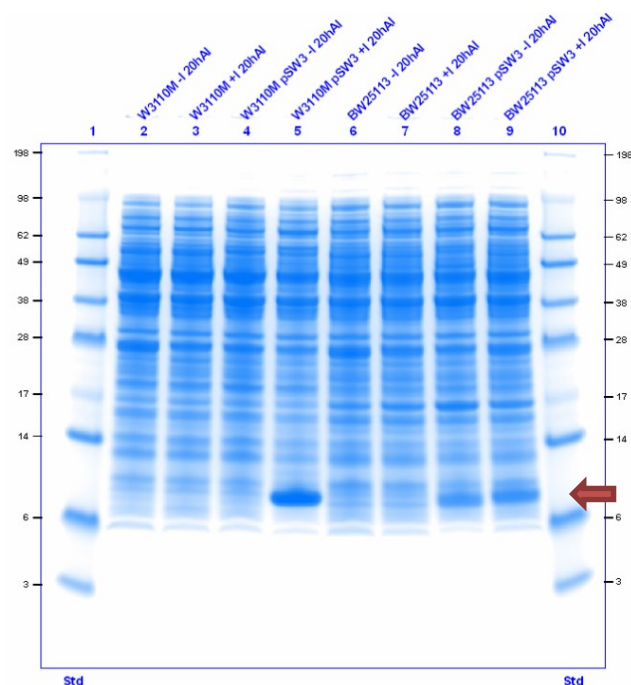


Figure 12. SDS-PAGE showing intracellular protein profile for different *E. coli* strains (*E. coli* K-12 W3110M, *E. coli* K-12 W3110M pSW3, *E. coli* K-12 BW25113 and *E. coli* K-12 BW25113 pSW3), grown on M9 medium, about 20h after induction of protein expression with IPTG (20hAI). Intracellular protein profiles of all tested strains and conditions correspond to an OD_{600nm} of 0.5, so that gel lanes are comparable. Plasmid pSW3 expresses an 11 kDa recombinant mini-proinsulin, indicated by the red arrow. For each tested strain, an IPTG-induced sample (+I) as well as a non-induced sample (-I) is included. Std refers to the SeeBlue® Plus2 Pre-stained Protein Standard. Molecular weights of the protein standard are shown in kDa.

As expected, for the control empty strains *E. coli* K-12 W3110M and *E. coli* K-12 BW25113 (not containing plasmid pSW3) no mini-proinsulin expression was reported for both induced and non-induced samples (Figure 12: lanes 2-3 and 6-7, respectively). Strain *E. coli* K-12 W3110M pSW3 showed good levels of mini-proinsulin expression 20 h after IPTG induction as well as really low promoter leakiness, since no mini-proinsulin expression was reported for the non-induced sample (Figure 12: lanes 4 and 5). However, as opposed to *E. coli* K-12 W3110M, P_{tac} promoter controlling expression of mini-proinsulin in plasmid pSW3 was leaky in the genetic background of *E. coli* K-12 BW25113, since there was no significant difference in reported mini-proinsulin expression levels between the induced and non-induced samples (Figure 12: lanes 8 and 9). In addition, expression intensity achieved by strain *E. coli* K-12 BW25113 pSW3 was clearly weaker than the one obtained for strain *E. coli* K-12 W3110M pSW3 (Figure 12: lanes 5 and 9). Hence, alternative pSW3 variants expressing different levels of LacI repressor were constructed and tested in the genetic background of *E. coli* K-12 BW25113 in order to avoid promoter leakiness and improve recombinant protein production.

5.2 Analysis of mini-proinsulin expression in *E. coli* K-12 BW25113 containing different variants of plasmid pSW3

The aim of this experiment was to compare the mini-proinsulin expression level as well as the leakiness of P_{tac} promoter by SDS-PAGE in cultures of *E. coli* K-12 BW25113 containing 3 different variants of plasmid pSW3 (pSW3, pSW3_*lacI*⁺ and pSW3_*lacI*^q), each of them expressing different amounts of LacI repressor. Strain *E. coli* K-12 W3110M expressing pSW3 was used as a reference for comparison since it previously showed optimal expression levels for recombinant mini-proinsulin (see section 5.1.2).

5.2.1 Cultivation conditions

Pre-culture was prepared as follows: 20 mL of modified Davis and Mingioli medium were introduced in sterile 100 mL Erlenmeyer shake flasks. One flask was prepared for each tested strain (*E. coli* K-12 W3110M pSW3, *E. coli* K-12 BW25113 pSW3, *E. coli* K-12 BW25113 pSW3_*lacI*⁺ and *E. coli* K-12 BW25113 pSW3_*lacI*^q). Ampicillin was supplemented to a final concentration of 100 µg/mL. Fresh cells were inoculated into the medium and cultures were incubated at 37 °C und 220 rpm, overnight. The main culture was prepared and cultivated as described as follows: 8 sterile 300 mL Erlenmeyer shake flasks were filled with 30 mL of modified Davis and Mingioli medium. 2 shake flasks were prepared for each tested strain. OD_{600nm} of pre-cultures was measured and the volume from the pre-culture needed to generate a main culture at 0.1 OD_{600nm} in a final volume of 30 mL was calculated. The corresponding calculated volume of pre-culture needed was introduced in the 300 mL sterile Erlenmeyer shake flasks. Ampicillin was supplemented to a final concentration of 100 µg/mL. Resulting cultures were incubated at 37°C and 220 rpm until OD_{600nm} reached 0.3-0.4. Then, 300 µL of 40% glucose were added to all cultures while IPTG was added to a concentration of 0.5 mM only to one of the shake flasks available for each tested strain (induced sample). The other shake flask was used as negative control for induction (non-induced sample). Cultures were then left incubating at 37°C and 220 rpm. Short before induction and 1h, 3h, 6h and overnight after induction, OD_{600nm} was measured and the corresponding volume needed to generate a 1 mL solution at OD_{600nm} =1 was taken out, centrifuged (10 000 g, 5 min) and the resulting pellet was stored at -20°C for SDS-PAGE analysis.

5.2.2 Mini-proinsulin analysis with SDS-PAGE

For each tested strain, mini-proinsulin expression was evaluated by SDS-PAGE with samples taken over cultivation time (1h, 3h, 6h and overnight after induction). In addition, for each strain, 2 cultivations were evaluated: one with IPTG induction and the other, without induction (Figure 13). Calculation of the relative quantity (Rel. q.) of mini-proinsulin was performed with the software Image Lab (Biorad) as follows: intensity of the band containing mini-proinsulin (band around 11 kDa present in induced samples in the SDS-PAGE gels) was determined for each lane. Intensity of the first band containing mini-proinsulin (band around 11 kDa present in lane 3 of the SDS-PAGE gels) was used as a reference for the calculation of the relative quantity of mini-proinsulin in the other bands.

For instance, the relative quantity of mini-proinsulin in a certain lane would be equal to the intensity of the mini-proinsulin band in such lane divided by the intensity of that band in lane 3.

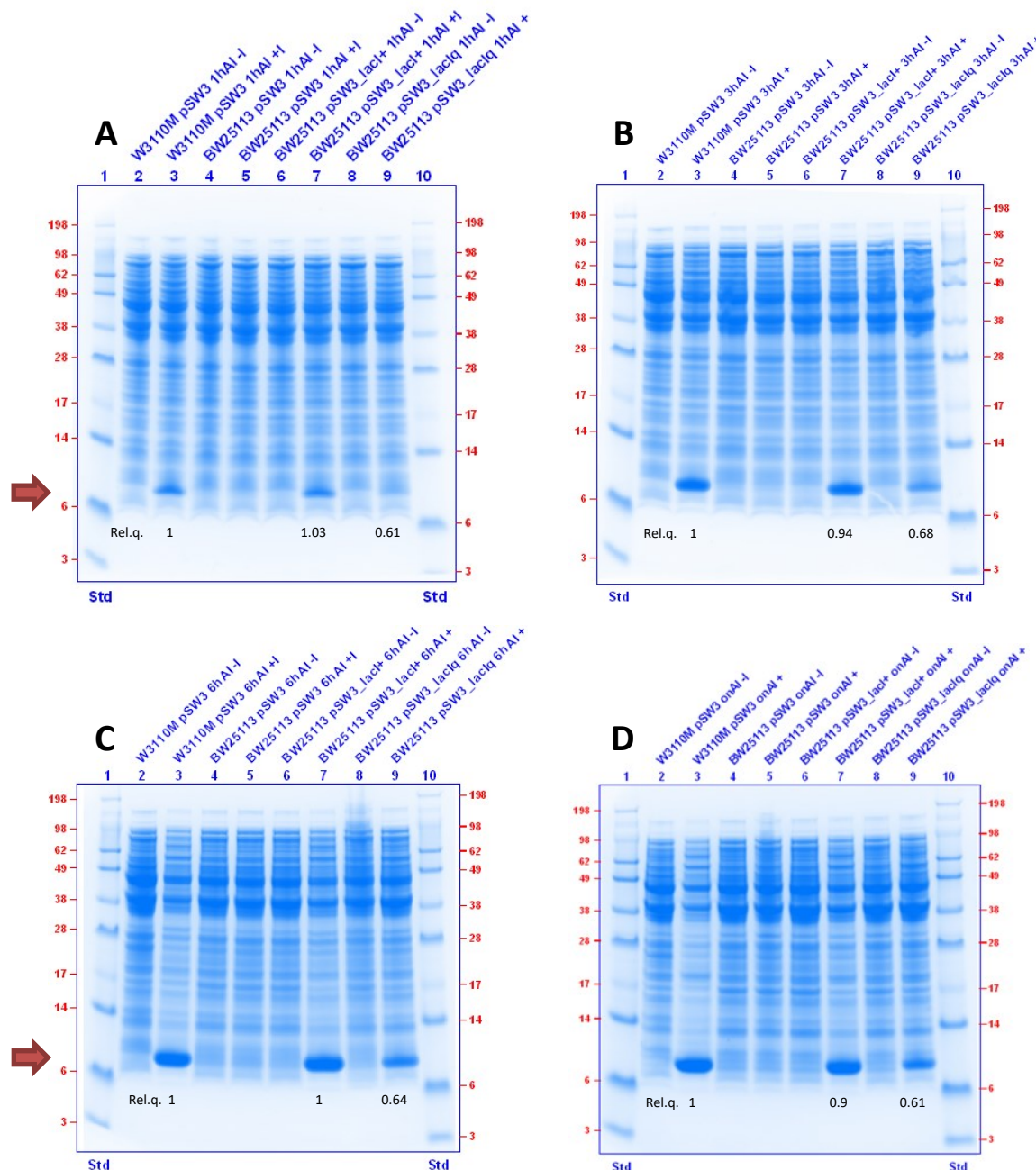


Figure 13. SDS-PAGE showing intracellular protein profile for different *E. coli* strains (*E. coli* K-12 W3110M pSW3, *E. coli* K-12 BW25113 pSW3, *E. coli* K-12 BW25113 pSW3_{lacI}⁺ and *E. coli* K-12 BW25113 pSW3_{lacI}^q), grown on modified Davis & Mingioli medium, 1h (A), 3h (B), 6h (C) and overnight (D) after induction of protein expression with IPTG (onAI). Intracellular protein profiles of all tested strains and conditions correspond to an OD_{600nm} of 0.5, so that gel lanes are comparable. Plasmid pSW3 and its variants express an 11 kDa recombinant mini-proinsulin, indicated by the red arrow. For each tested strain, an IPTG-induced sample (+I) as well as a non-induced sample (-I) is included. Std refers to the SeeBlue® Plus2 Pre-stained Protein Standard. Molecular weights of the protein standard are shown in kDa. Relative quantity (Rel. q.) of mini-proinsulin is also indicated.

According to the results, *E. coli* K-12 BW25113 (*lacI*⁺) expressing pSW3_{lacI}⁺ presents a similar induction behavior than *E. coli* K-12 W3110M (*lacI*^q) expressing pSW3 (Figure 13: lanes 2-3 and 6-7). *E. coli* K-12 BW25113 expressing pSW3 did not show any expression after IPTG induction (Figure 13:

lane 5). However, in section 5.1.2, protein production was reported for that strain 20 h after induction. This incongruity is explained due to an experimental error, since strain employed in this study was not actually *E. coli* K-12 BW25113 expressing pSW3 but the empty *E. coli* K-12 BW25113 strain. Strain *E. coli* K-12 BW25113 (*lacI*⁺) expressing pSW3_*lacI*^q also achieved protein expression (Figure 13: lane 9). However, expression reported around a 40 % decrease than in *E. coli* K-12 W3110M pSW3 and *E. coli* K-12 BW25113 pSW3_*lacI*⁺ (Figure 13: lanes 3, 7 and 9). In conclusion, it seems that for the genetic background of *E. coli* K-12 BW25113 (*lacI*⁺) the pSW3 plasmid variant containing the *lacI*⁺ genetic region is the optimal one since it shown a similar induction behavior than *E. coli* K-12 W3110M (*lacI*^q) pSW3, thus achieving a good production of mini-proinsulin. In addition, P_{tac} promoter seems to be quite good regulated since, in the absence of IPTG, no production was achieved (Figure 13: lane 6). Hence, strain *E. coli* K-12 BW25113 (*lacI*⁺) pSW3_*lacI*⁺ was selected as the model organism for further investigation in the current thesis.

5.3 Evaluation of L-arabinose induction in *E. coli* BW25113 Δ *geneX* expressing pSW3_*lacI*⁺ and pACG_*araBAD_geneX* (*geneX*-tunable *E. coli*)

The aim of this experiment was to evaluate the effect of using different concentrations of L-arabinose in the expression level of target genes involved in the BCAA biosynthetic pathway (*gene X*), which are under the control of the *araBAD* promoter in strains *E. coli* K-12 BW25113 Δ *geneX* expressing pSW3_*lacI*⁺ and pACG_*araBAD_geneX* (*geneX*-tunable *E. coli*).

5.3.1 Cultivation conditions

Pre-cultures were prepared as follows: 5 mL of 1:3 TUB medium were introduced in sterile 15 mL tubes. Various pre-cultures were prepared for each tested strain. Ampicillin and chloramphenicol were supplemented to a final concentration of 100 µg/mL and 25 µg/mL, respectively. Chloramphenicol was not added in cultures containing strain *E. coli* K-12 BW25113 pSW3_*lacI*⁺. Different concentrations of L-arabinose were added to the different pre-cultures. 50 µL of the corresponding criostock were then inoculated into the medium and cultures were incubated at 37 °C and 250 rpm, overnight. OD_{600nm} was measured at 16h cultivation time.

5.3.2 Gene expression analysis

Most of the target genes investigated in this study are crucial for *E. coli* metabolism so that when no L-arabinose is added, target gene is not expressed and this might have a negative effect on *E. coli* growth behavior. Hence, in order to indirectly proof effectivity of L-arabinose as inducer of gene expression in plasmids pACG_*araBAD_geneX*, OD_{600nm} from the different pre-cultures was measured after 16h cultivation. Growth behavior (OD_{600nm}) was then plotted for every tested tunable *E. coli* strain under different L-arabinose concentrations and it was compared to the non-modified strain *E. coli* K-12 BW25113 pSW3_*lacI*⁺ (

Figure 14).

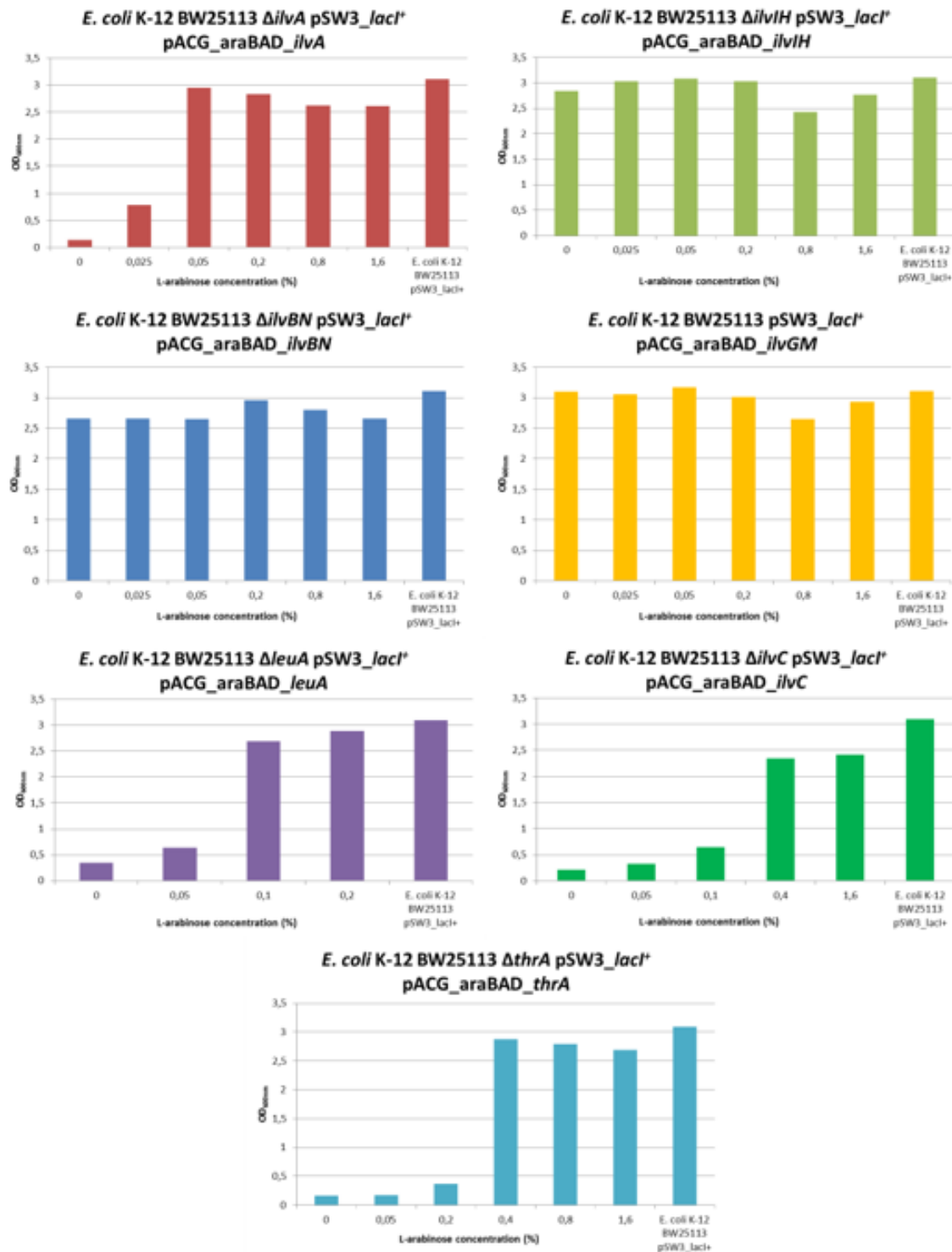


Figure 14. Chart showing measured OD_{600nm} after 16h cultivation of the different *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_{lacI}⁺ pACG_araBAD_{geneX} strains under different L-arabinose concentrations. The non-modified strain *E. coli* K-12 BW25113 pSW3_{lacI}⁺ is also included as a control for comparison.

For strain *E. coli* K-12 BW25113 $\Delta ilvA$ pSW3_{lacI}⁺ pACG_araBAD_{ilvA} (simplified, *ilvA*-tunable *E. coli*), L-arabinose induction showed a positive growth effect when using at least 0.025 % L-arabinose. Growth behavior recovered levels of the control *E. coli* strain (*E. coli* K-12 BW25113 pSW3_{lacI}⁺) when at least 0.05 % L-arabinose was employed. For *ilvIH*-, *ilvBN*- and *ilvGM*-tunable *E. coli* strains

there were no significant differences in growth behavior reported between non-induced and L-arabinose-induced samples, independently on the L-arabinose concentration employed. In those cases, induction efficiency of L-arabinose could not be properly evaluated by monitoring cell growth. L-arabinose induction showed a positive growth effect for *leuA*-tunable *E. coli* when using at least 0.05 % L-arabinose. Growth behavior recovered levels of the control *E. coli* strain when at least 0.1 % L-arabinose was employed. L-arabinose was also proven to function for *ilvC*-tunable *E. coli*. A positive growth effect was reported when using at least 0.05 % L-arabinose. Growth behavior approached to levels of the control *E. coli* strain when at least 0.4 % L-arabinose was employed. Finally, L-arabinose induction was also demonstrated to success for *thrA*-tunable *E. coli*. When employing at least 0.4 % L-arabinose a positive growth effect was reported, approaching to levels of the control *E. coli* strain.

It could be concluded that the effective induction range of L-arabinose for plasmid variants pACG_araBAD_ *geneX* was between 0.05 and 1.6 % L-arabinose. Furthermore, the maximum tested L-arabinose concentration (1.6 %) did not show a cellular toxic effect, since growth behavior remained unaltered. It is also noteworthy that, for each target gene, different induction strength, i.e. L-arabinose concentration, was necessary in order to trigger genetic expression levels enough to recover cell growth levels of the control *E. coli* strain.

5.4 Evaluation of m-toluate induction in *E. coli* BW25113 Δ *geneX* expressing pSW3_ *lacI*⁺ and pACG_XylSPm_ *geneX* (*geneX*-tunable *E. coli*)

Analog to section 5.3, the aim of this experiment was to evaluate the effect of employing different concentrations of m-toluate in the expression of target genes involved in the BCAA biosynthetic pathway (*gene X*), which are under the control of the XylS/Pm promoter in strains *E. coli* K-12 BW25113 Δ *geneX* expressing pSW3_ *lacI*⁺ and pACG_XylSPm_ *geneX* (*geneX*-tunable *E. coli*).

5.4.1 Cultivation conditions

The cultivation procedure was as described in section 5.3.1, with the following alterations: instead of 1:3 TUB medium, modified Davis & Mingioli medium was employed; instead of L-arabinose, different concentrations of m-toluate were added to the pre-cultures.

5.4.2 Gene expression analysis

Similarly to section 5.3.2, OD_{600nm} was plotted for every tested tunable *E. coli* strain under different m-toluate concentrations and it was compared to the non-modified strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ (Figure 15).

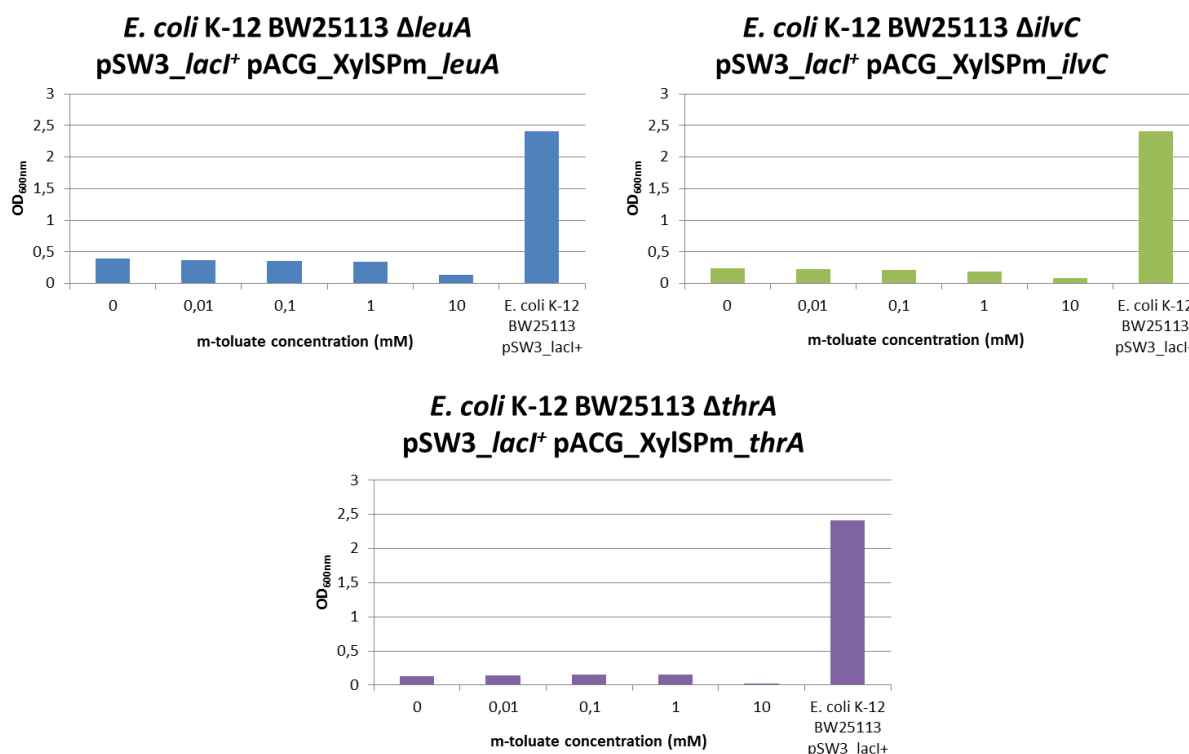


Figure 15. Chart showing measured OD_{600nm} after 16h cultivation of the different *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_lacI⁺ pACG_XylSPm_geneX strains under different L-arabinose concentrations. The non-modified strain *E. coli* K-12 BW25113 pSW3_lacI⁺ is also included as a control for comparison.

Induction of plasmid variants pACG_XylSPm_geneX with m-toluate was not successful since addition of m-toluate did not show a positive effect on growth behavior if compared with the non-induced case for any tested mutant strain. In addition, it can be concluded that 10 mM m-toluic acid triggers cellular toxicity since under this concentration OD_{600nm} reached almost 0 for all tested strains.

In accordance to results obtained in sections 5.3.2 and 5.4.2, the plasmid variant pACG_araBAD_geneX was selected in this thesis for further expression regulation of target genes involved in the BCAA biosynthetic pathway. The plasmid variant pACG_XylSPm_geneX was discarded.

5.5 Establishment of a GC-FID method allowing analysis of canonical and non-canonical amino acids

The aim of this experiment was to establish and validate the use of the EZ:faast™ free (physiological) amino acid analysis kit for canonical and non-canonical amino acid analysis by gas chromatography-flame ionization detection (GC-FID). To achieve that, following activities were carried out: determination of retention times and elaboration of calibration curves for each amino acid (sections 5.5.1 and 5.5.2), evaluation of the hydrolysis effect on amino acid analysis (section 5.5.3) and validation of the method by analyzing a pure protein (section 5.5.4).

5.5.1 Elaboration of calibration curves and determination of retention times for ncBCAA norvaline, norleucine and β -methylnorleucine

The following amino acid mixtures were prepared for calibration: calibration level 1 (25 μ M): 12.5 μ L 200 μ M ncBCAA plus 100 μ L 200 μ M ABA; calibration level 2 (50 μ M): 25 μ L 200 μ M ncBCAA plus 100 μ L 200 μ M ABA; calibration level 3 (100 μ M): 50 μ L 200 μ M ncBCAA plus 100 μ L 200 μ M ABA; calibration level 4 (200 μ M): 100 μ L 200 μ M ncBCAA plus 100 μ L 200 μ M ABA. Each mixture was used for amino acid analysis according to sections 4.8.3-4.8.5. Retention times for each ncBCAA were determined (Table 24) and calibration curves were elaborated by plotting the ratio of the analyte signal (ncBCAA) to the internal standard signal (ABA) as a function of the analyte concentration (Figure 16).

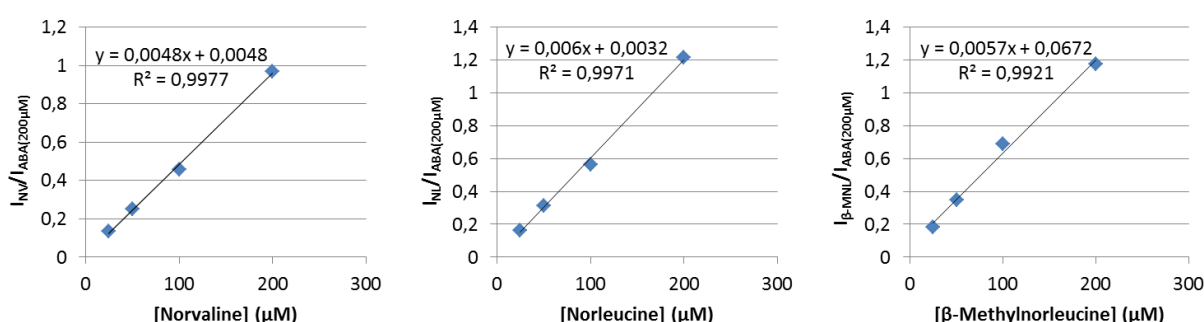


Figure 16. Calibration curves for ncBCAA. The y-axis in the calibration curves represents the ratio of the area of a given peak corresponding to one amino acid by the peak area of ABA ($I_{ncBCAA}/I_{ABA(200 \mu M)}$). The x-axis refers to the real amino acid concentration. Coefficient of determination (R^2) and linear regression equation are also shown.

5.5.2 Elaboration of calibration curves and determination of retention times for canonical amino acids

The following amino acid mixtures were prepared for calibration: calibration level 1 (50 μ M): 25 μ L 200 μ M SD1 or SD2 plus 100 μ L 200 μ M ABA; calibration level 2 (100 μ M): 50 μ L 200 μ M SD1 or SD2 plus 100 μ L 200 μ M ABA; calibration level 3 (200 μ M): 100 μ L 200 μ M SD1 or SD2 plus 100 μ L 200 μ M ABA; calibration level 4 (400 μ M): 200 μ L 200 μ M SD1 or SD2 plus 100 μ L 200 μ M ABA. Each mixture was used for amino acid analysis according to sections 4.8.3-4.8.5. Retention times for each canonical amino acid were determined (Table 24) and calibration curves were elaborated by plotting the ratio of the analyte signal (canonical amino acid) to the internal standard signal (ABA) as a function of the analyte concentration (Figure 17).

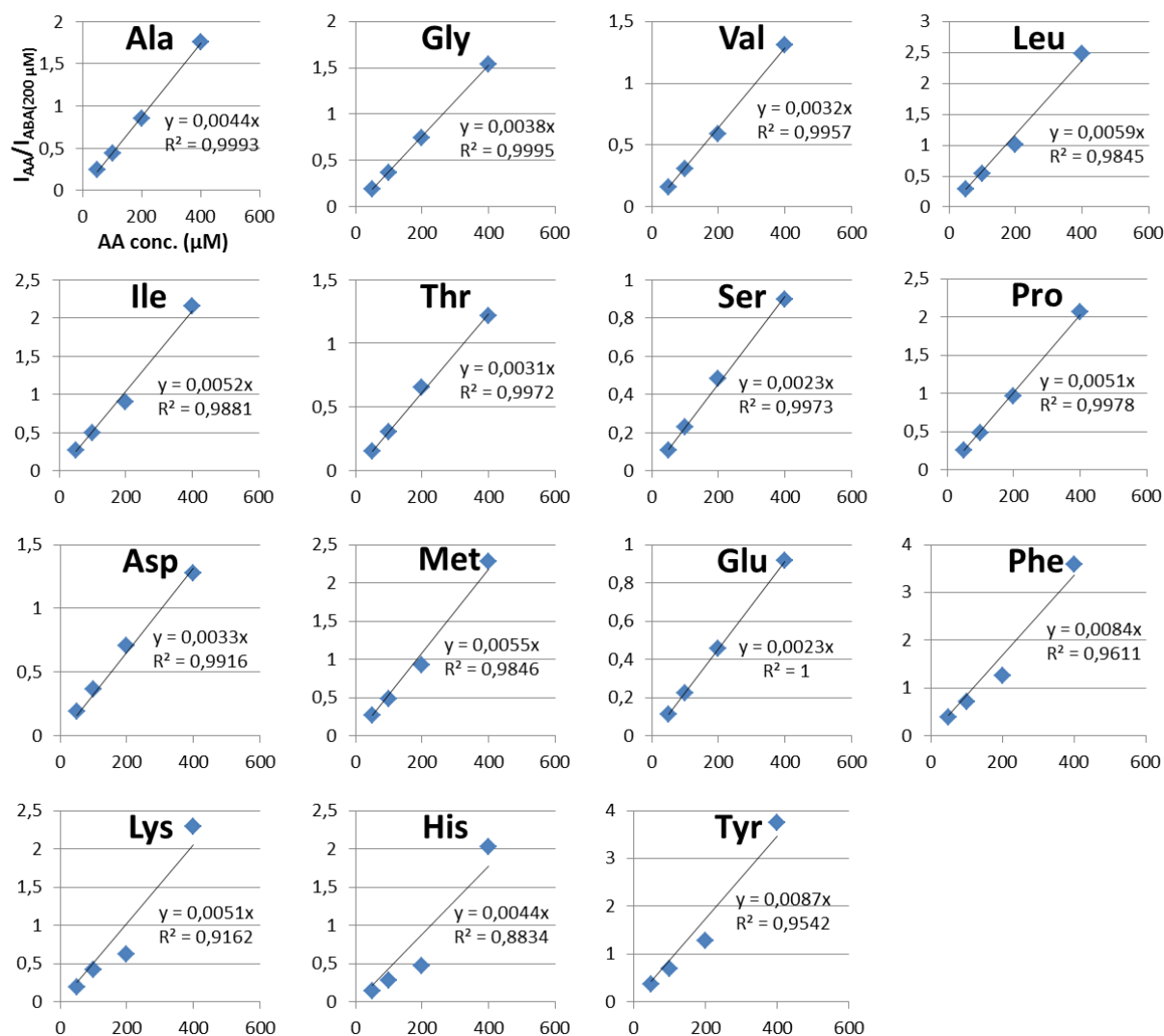


Figure 17. Calibration curves for canonical amino acids. The y-axis represents the ratio of the area of a given peak corresponding to one amino acid by the peak area of ABA ($I_{AA}/I_{ABA(200 \mu M)}$). The x-axis refers to the real amino acid concentration in μM units. Coefficient of determination (R^2) and linear regression equation are also shown.

Table 24. Retention times for each tested amino acid.

Amino acid	Retention time (min)	Amino acid	Retention time (min)
alanine	2.461	threonine	3.447
glycine	2.601	serine	3.517
2-aminobutyric acid	2.728	proline	4.190
valine	2.839	aspartate	4.210
norvaline	3.099	methionine	4.550
leucine	3.158	glutamate	4.584
isoleucine	3.282	phenylalanine	5.843
norleucine	3.340	lysine	6.033
β -methylnorleucine	3.398 and 3.447	histidine	6.311
		tyrosine	6.307

5.5.3 Evaluation of the effect of hydrolysis on amino acid analysis

For amino acid analysis of protein samples, a hydrolysis approach has to be previously followed in order to obtain the free amino acids. The most employed hydrolysis method is the acid hydrolysis with 6N HCl. However, according to literature (Pickering and Newton, 1990; Davidson, 1997), the use of acid hydrolysis can have various negative impacts affecting subsequent amino acid analysis, which needed to be evaluated in this work. In order to assess the effects of acid hydrolysis, amino acid content of standard solutions, being or not subjected to the hydrolysis process, was analyzed by GC-FID and results were compared. The standard solutions consisted of 200 μ M ncBCAA, 200 μ M SD1 and 200 μ M SD2. The standard solution ncBCAA contained amino acids norvaline, norleucine and β -methylnorleucine, the standard solution SD2 comprised amino acids asparagine, glutamine and tryptophan and standard solution SD1 contained the remaining canonical amino acids.

For samples subjected to hydrolysis, the hydrolysis protocol was applied as described in section 4.8.2. In the current case, 200 μ L of the corresponding standard solution were mixed with 800 μ L 5M HCl. The resulting hydrolyzed sample pellets were then resuspended with 200 μ L of a solution containing 20 mM HCl and 10% isopropanol. 100 μ L of the resulting suspension were then mixed with 100 μ L of 200 μ M ABA and these solutions were used for amino acid analysis according to described in sections 4.8.3-4.8.5. For samples not subjected to hydrolysis, protocol applied was as follows: 100 μ L of the corresponding standard solution were mixed with 100 μ L of 200 μ M ABA and resulting solutions were used for amino acid analysis according to described in sections 4.8.3-4.8.5. Amino acid content from both solution sets was analyzed by GC-FID and the % of variation of the determined amino acid concentration of hydrolyzed amino acid samples with respect to non-hydrolyzed samples was calculated for each amino acid (Figure 18).

For most of the amino acids, a signal reduction was reported after hydrolysis, being this especially evident for norvaline, norleucine, threonine, methionine and tyrosine. Interestingly, the signal of glutamate reported an increase after hydrolysis. In addition, asparagine and glutamine were respectively deamidated to aspartate and glutamate after hydrolysis, while tryptophan was destroyed.

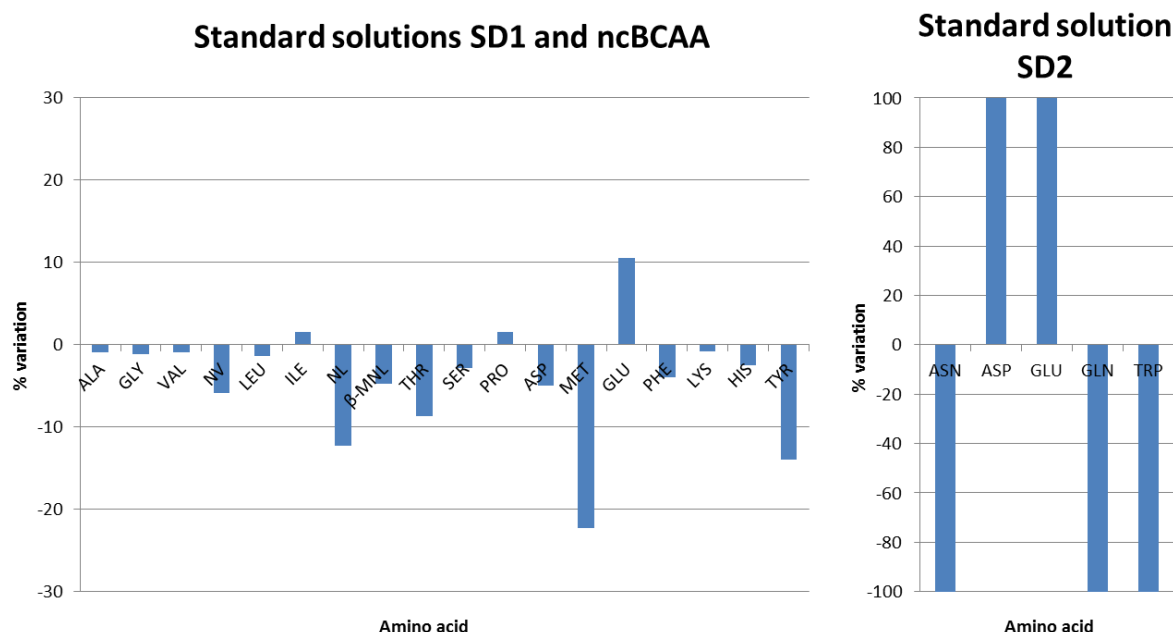


Figure 18. Charts representing the % of variation of the determined amino acid concentration of hydrolyzed amino acid samples with respect to non-hydrolyzed samples for each amino acid. The chart in the left presents the % of variation of amino acids present in standard solutions SD1 and ncBCAA while the chart in the right shows the % of variation of amino acids present in standard solution SD2.

5.5.4 Validation of the GC-FID method by analyzing a pure protein

In order to validate the established GC-FID method, a pure pre-proinsulin (PPI) variant, whose amino acid composition and molecular weight are known, was employed for amino acid analysis by GC-FID. Solution containing pure PPI was hydrolyzed as described in section 4.8.2. In the current case, 250 μL of a 10 mg/mL PPI solution were mixed with 750 μL 5M HCl. Hydrolysis was performed at both 24 h and 72 h in order to identify if hydrolysis time has an effect on amino acid analysis. The hydrolyzed protein pellet was resuspended with 250 μL dH_2O . Protein concentration of the resulting solution was measured by Nanodrop and amino acid concentration was then calculated according to

following equation:
$$\text{Amino acid concentration} \left(\frac{\mu\text{mol}}{\mu\text{L}} \right) = \frac{\text{Protein concentration} \left(\frac{\text{mg}}{\text{mL}} \right) \cdot \text{moles amino acid}}{\text{Protein molecular weight (kDa)} \cdot 10^3}.$$

The volume of hydrolyzed PPI solution corresponding to 10 μmol amino acids was mixed with 100 μL 200 μM ABA. The resulting samples were treated according to described in sections 4.8.3-4.8.5 and amino acid concentrations were determined. From that data, the experimental amino acid composition of the PPI was calculated and this was compared to the theoretical amino acid composition of PPI (Figure 19). Alanine was used as the reference amino acid for those calculations since, according to results obtained at the previous section 5.5.3, alanine was the amino acid less affected by hydrolysis.

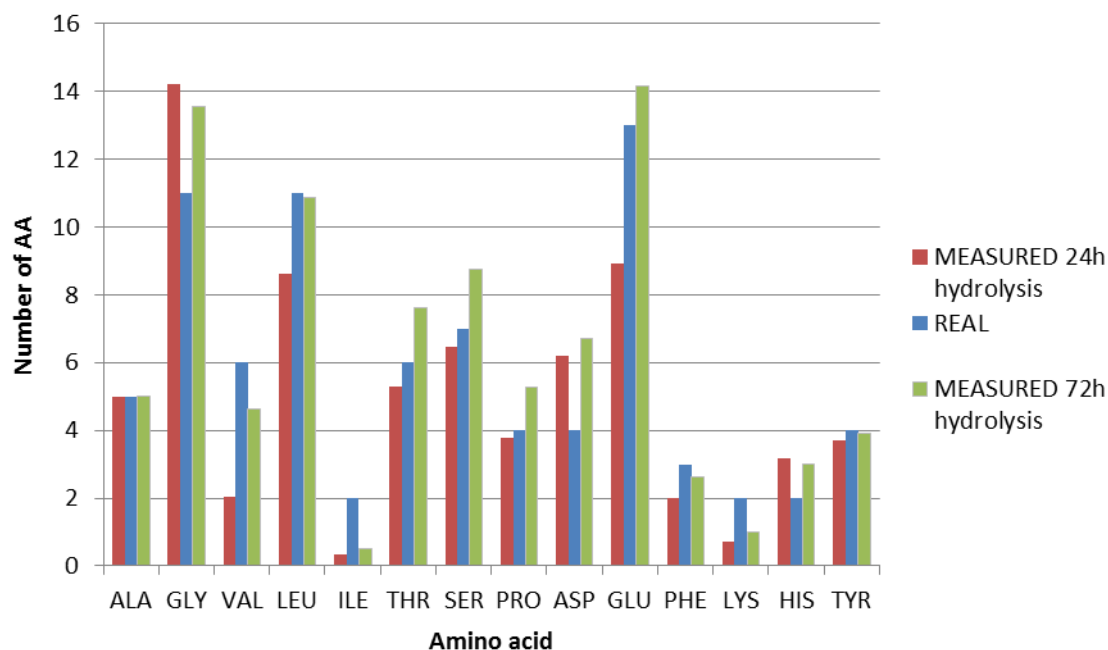


Figure 19. Comparison of the theoretical amino acid composition of PPI (blue) with the experimental amino acid composition of PPI calculated after GC-FID analysis of pure PPI samples subjected to 24h (red) or 72h (green) acid hydrolysis.

Experimental results matched quite accurately the theoretical amino acid composition. Moreover, 72 h hydrolysis slightly improved results: leucine concentration increased up to theoretical values; threonine, serine, proline and glutamate concentrations also increased, reaching values slightly above theoretical values; valine, isoleucine, phenylalanine and lysine concentrations increased as well, but they still remained below real values. However, since 72 h hydrolysis requires 2 additional days and the improvement of hydrolysis performance is not so dramatic, 24 h hydrolysis was selected for further experimentation.

5.6 Establishment of cultivation conditions based on pyruvate pulsing leading to an increase of ncBCAA mis-incorporation into recombinant mini-proinsulin expressed in *E. coli* at shake flask level

The aim of this experiment was to mimick large-scale effects, i.e. increase of ncBCAA biosynthesis and minisincorporation into recombinant mini-proinsulin, at shake flask level by exposing an *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation to pyruvate pulses of different intensity.

5.6.1 Cultivation conditions

Pre-culture was prepared as follows: 20 mL of TUB medium were introduced in a sterile 100 mL Erlenmeyer shake flask. Ampicillin was supplemented to a final concentration of 100 µg/mL. Fresh cells from strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ were inoculated into the medium and culture was incubated at 37 °C and 220 rpm, overnight.

The main culture was prepared and cultivated as follows: 7 sterile 300 mL Erlenmeyer shake flasks were filled with 50 mL TUB medium. OD_{600nm} of the pre-culture was measured and the volume from the pre-culture needed to generate a main culture at 0.1 OD_{600nm} in a final volume of 50 mL was calculated. The corresponding calculated volume of pre-culture was introduced in the 300 mL sterile Erlenmeyer shake flasks. Ampicillin was supplemented to a final concentration of 100 µg/mL. Resulting cultures were incubated at 37°C and 220 rpm. When OD_{600nm} reached 0.3, pulses of different pyruvate concentrations were applied each 30 minutes for a time period of 2.5h (a total of 6 pyruvate pulses). Seven different pyruvate concentrations were tested, one for each available shake flask: 0, 100, 250, 500, 1000, 2000 and 5000 mg/L. Induction with 0.5 mM IPTG was performed when cultures reached OD_{600nm} of 0.6. Around 3.5h after induction cultivations were stopped, OD_{600nm} was measured and cells contained in the whole liquid culture were harvested by centrifugation at 10000 g for 10 min for further amino acid analysis by GC-FID.

5.6.2 Analysis of ncBCAA

Intracellular soluble protein fraction and inclusion body isolation from cell extracts was carried out according to described in section 4.8.1. Acid hydrolysis of intracellular soluble and inclusion body protein fractions was performed as described in section 4.8.2. In the current case, 250 µL of the intracellular soluble protein fraction were mixed with 750 µL 5M HCl. Isolated inclusion body pellets were resuspended with 200 µL H₂O and afterwards 750 µL 5M HCl were added. Hydrolyzed IB pellet samples were resuspended with 500 µL dH₂O while hydrolyzed intracellular soluble protein fraction samples were resuspended with 1 mL dH₂O. The resulting hydrolyzed samples were treated according to described in section 4.8.3. In the current case, 50 µL of the hydrolyzed intracellular soluble protein fraction or the hydrolyzed inclusion body fraction were mixed with 100 µL of 200 µM ABA. The resulting samples were used for amino acid analysis by GC-FID according to the procedure described in sections 4.8.4-4.8.5. Concentrations of ncBCAA in both inclusion body and intracellular soluble fractions under the different tested cultivation conditions are shown in Figure 20.

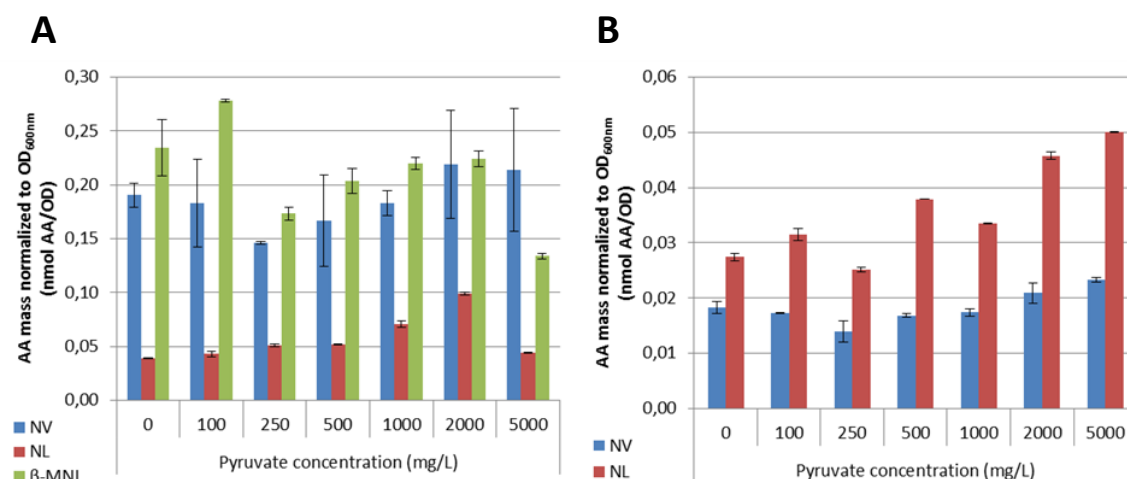


Figure 20. Molar concentrations of norleucine (red bars), norvaline (blue bars) and β -methylnorleucine (green bars) normalized to OD_{600nm} present in the intracellular soluble protein fraction (A) and inclusion body fraction (B) from samples taken 3.5 h after IPTG induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivations in shake flasks under cultivation conditions subjected to pulses of different pyruvate concentration. When cultivation achieved OD_{600nm} of 0.3, pyruvate was supplemented in a pulse-based manner each 30 minutes for a time period of 2.5h, this is a total of 6 pyruvate pulses. Depending on the cultivation pyruvate pulses were added up to 0, 100, 250, 500, 1000, 2000 and 5000 mg/L. Induction with 0.5 mM IPTG was performed when cultures reached OD_{600nm} of 0.6.

For both intracellular soluble protein and inclusion body fractions, addition of increasing concentrations of pyruvate triggered norleucine formation. However, it is noteworthy to point out that norleucine concentration calculated in the intracellular soluble fraction under 5 g/L pyruvate was clearly out of the general trend. This value was considered as an outlier and not employed for further data interpretation. The positive effect of pyruvate in norvaline formation was less evident. In the inclusion body fraction, a positive trend was shown from 250 mg/L to 5 g/L pyruvate. However, in the intracellular soluble fraction a clear trend could not be observed due to the high variance. This observation also applied for β -methylnorleucine (Figure 20). Similar results were obtained when expressing data as ratio of the concentration of ncBCAA with respect to the canonical counterpart (Figure S70).

5.7 Establishment of cultivation conditions based on pyruvate pulsing and O₂ limitation leading to an increase of ncBCAA mis-incorporation into recombinant mini-proinsulin expressed in *E. coli* in a 10 mL mini-reactor and in a 15L reactor

In section 5.6 it was proven that addition of pyruvate pulses effectively triggers an increase in norleucine biosynthesis and mis-incorporation when cultivating *E. coli* K-12 BW25113 expressing plasmid pSW3_ *lacI*⁺ at shake flask level. However, such effect was not clearly shown for the ncBCAA norvaline and β -methylnorleucine. It might be that cultivation conditions present at shake flask level were not optimal to clearly evaluate the effect of pyruvate on ncBCAA formation. Hence, the next step in this investigation was to perform similar experiments with pyruvate in more controlled cultivation systems such as mini-reactor and 15L reactor.

5.7.1 Experiment in a 10 mL mini-reactor

5.7.1.1 Cultivation mode

5.7.1.1.1 Standard cultivation

30 μ L of a cryostock containing *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ were used to inoculate 30 mL of 1:3 supplemented TUB mineral salt medium containing 5 g/L glucose, 0.1 M Na-Phosphate buffer and 100 μ g/mL ampicillin in order to generate the pre-culture. The pre-culture was incubated at 37 °C and 220 rpm in an orbital shaker, overnight. OD_{600nm} at the end of the pre-culture was measured and a given volume was used to inoculate a 5 mL-starting volume mini-reactor so that initial OD_{600nm} was 0.4. The mini-reactor medium consisted of a 1:3 supplemented TUB mineral salt medium containing 4 g/L glucose, 0.1 M Na-Phosphate buffer, 100 μ g/mL ampicillin and 1 μ L/mL antifoam Desmophen. Cultivation was carried out at 37 °C and the pH was maintained at 7 by automatic control with NH₄OH and CO₂. Stirrer speed was set to 800 rpm and DO set-point to 25 %, maintaining the last by automatically increasing the oxygen flow into the mini-reactor. Batch phase lasted around 4h. After batch phase was finished, 1 mL 400 g/L EnPump 200 solution and 50 μ L 3000 U/L amylase solution were manually added into the mini-reactor, hence starting the fed-batch phase. EnPump 200 is a glucose polymer and when amylase is present, it constantly hydrolyses the polymer, thus delivering free glucose molecules over time, ensuring then a glucose-limited fermentation. In order to generate the 400 g/L EnPump 200 solution, 25 g EnPump 200 powder was dissolved in a 25 mL solution, having the same composition than the medium already present in the mini-reactor, so that components' concentration in the mini-reactor remained invariable after adding the EnPump 200 solution. 30 min after beginning of the fed batch phase, expression of recombinant mini-proinsulin was induced by manual addition of an IPTG pulse to a final concentration of 0.5 mM. Fed-batch phase was active for 3.5 h. A general overview of the cultivation is shown in Figure 21. This cultivation was performed in triplicates in mini-reactor wells A1, A2 and A3.

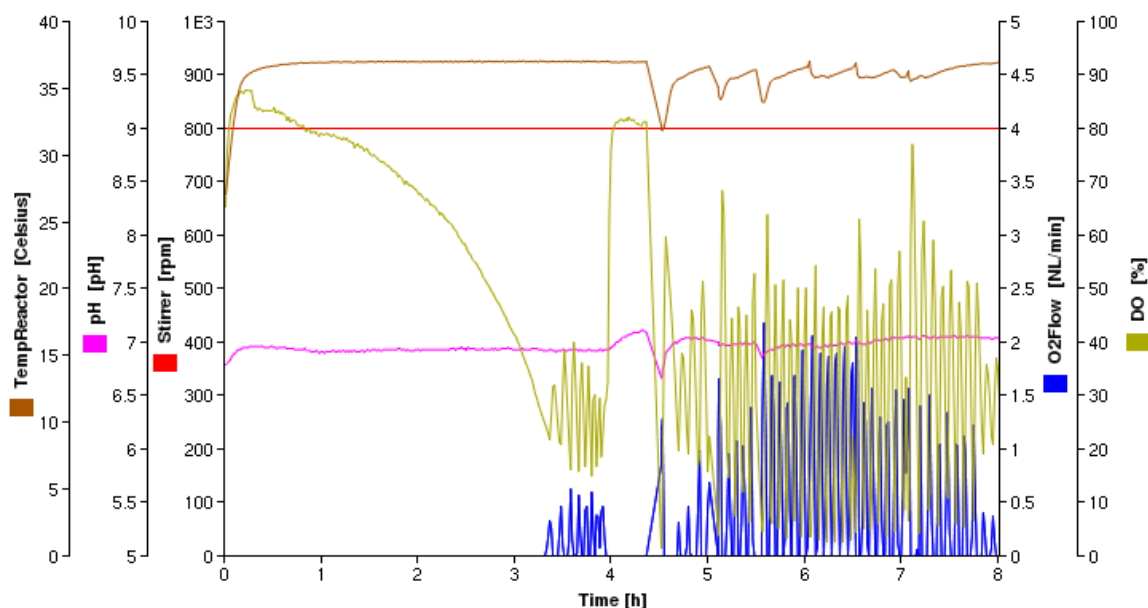


Figure 21. Overview of the standard cultivation of *E. coli* K-12 BW25113 pSW3_ *lacI** in a 10mL mini-reactor. A 4 h batch phase was followed by a 3.5 h EnPump-based fed- batch phase. Induction was carried out 30 min after fed-batch began (at 5h cultivation time). Cultivation was carried out at 37 °C and the pH was maintained at 7 by automatic control with NH₄OH and CO₂. Stirrer speed was set to 800 rpm and DO set-point to 25 %, controlling the last by automatically modifying the oxygen flow into the mini-reactor.

5.7.1.1.2 Cultivation triggering increase of ncBCAA production

Cultivation was performed as described for the standard cultivation in section 5.7.1.1.1. However, immediately after beginning of the fed-batch phase, a 0.833 g/L pyruvate pulse was manually added into the reactor. During the following 5 min after pyruvate addition, DO set-point was set to 0, so that no oxygen was supplied into the mini-reactor during that period, hence ensuring oxygen limitation. 30 min after the first pyruvate pulse, expression of recombinant mini-proinsulin was induced by manual addition of an IPTG pulse to a final concentration of 0.5 mM. After induction, sequential 0.833 g/L pyruvate pulses were manually performed each 30 min as described above for a total of 5 pulses. Between pulses, DO set-point was re-established to 25 %. Fed-batch phase was active for 3.5 h. A general overview of the cultivation is shown in Figure 22. This cultivation was performed in triplicates in mini-reactor wells A4, A5 and A6.

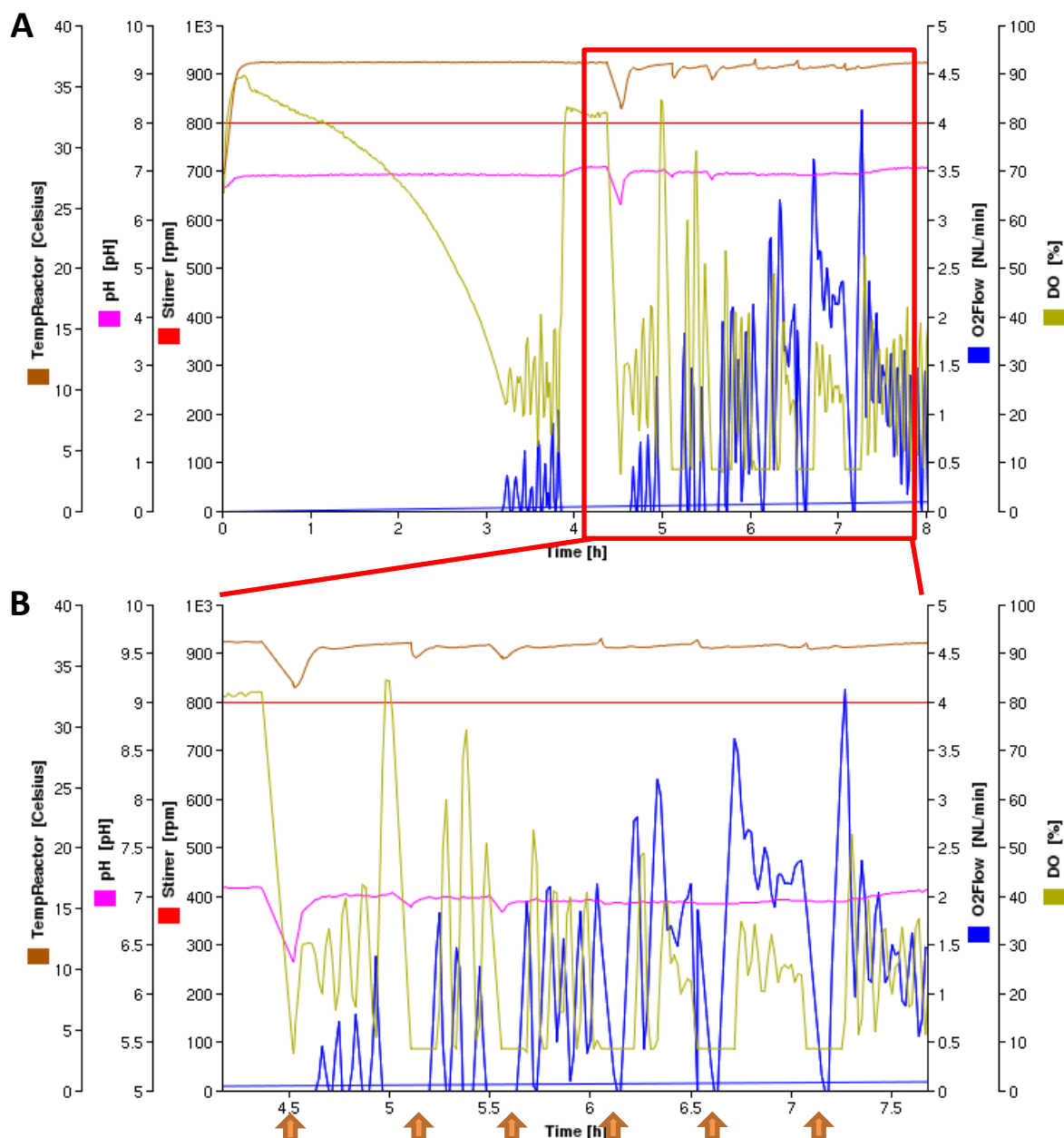


Figure 22. Overview of the cultivation of *E. coli* K-12 BW25113 pSW3_lacI* in a 10mL mini-reactor under conditions triggering increase of ncBCAA production, during the whole cultivation process (A) and during the fed-batch period (B). A 4 h batch phase was followed by a 3.5 h EnPump-based fed-batch phase. Induction was carried out 30 min after fed-batch began (at 5h cultivation time). Right after beginning of fed-batch phase and then each 30 min, a pyruvate pulse was added (indicated by orange arrows in B). 6 pyruvate pulses corresponding to 5 g/L pyruvate were added in total. During the following 5 min after pyruvate addition, DO set-point was set to 0, so that no oxygen was supplied into the mini-reactor. Cultivation was carried out at 37 °C and the pH was maintained at 7 by automatic control with NH₄OH and CO₂. Stirrer speed was set to 800 rpm and DO set-point to 25 %, controlling the last by automatically modifying the oxygen flow into the mini-reactor.

5.7.1.2 Mini-proinsulin analysis by SDS-PAGE from total cell extracts

Recombinant mini-proinsulin expression was evaluated by SDS-PAGE by densitometry analysis. Samples consisted on OD1/mL cell pellets and were taken 3h after induction. Samples were processed for SDS-PAGE analysis according to described in section 4.7.15. Scanned 2-dimensional (2D) SDS-PAGE gels are shown in Figure 23, where each gel lane corresponds to a cell suspension of $OD_{600nm}=0.25$. In order to allow quantification of mini-proinsulin content in the samples, a pure pre-proinsulin (PPI) calibrate was employed. Hence, 1 μ g, 0.5 μ g and 0.25 μ g pure pre-proinsulin were loaded into each gel.

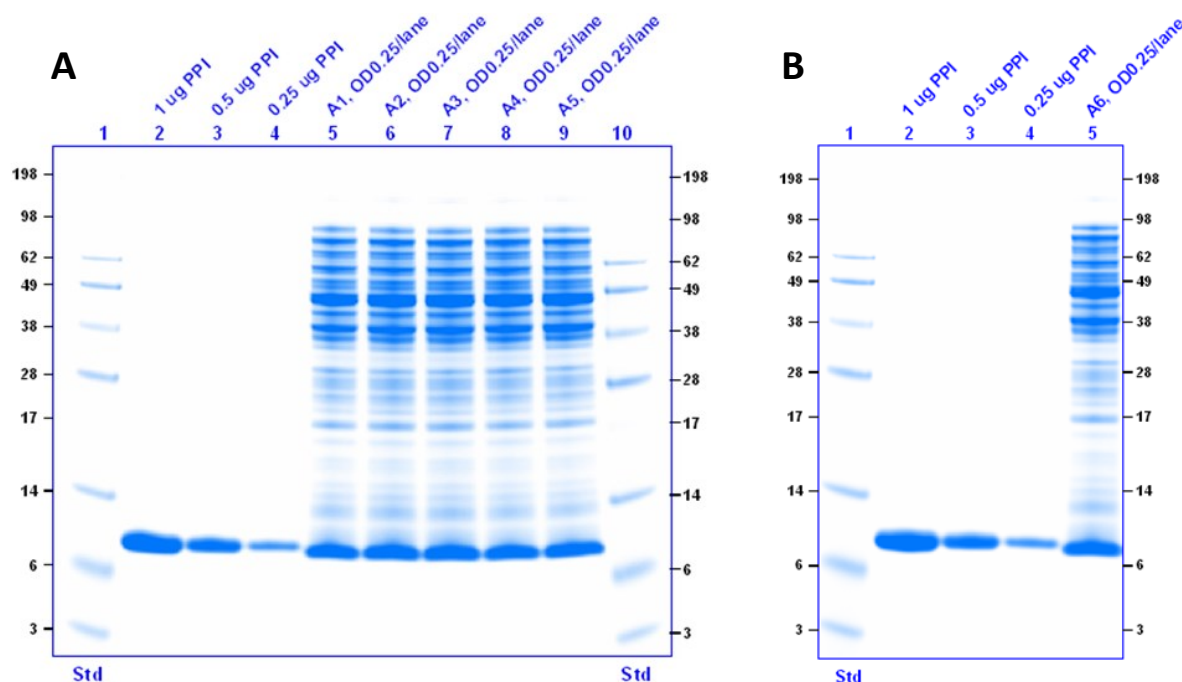


Figure 23. 2D SDS-PAGE gels of total cell extracts from samples A1, A2, A3, A4, A5 (A) and A6 (B). Each gel lane corresponds to a cell suspension of $OD_{600nm}=0.25$. In order to allow quantification of recombinant mini-proinsulin content in the samples by densitometry analysis, a pure pre-proinsulin having the same molecular weight than the expressed recombinant protein was employed. Hence, 1 μ g, 0.5 μ g and 0.25 μ g pure pre-proinsulin were loaded into each gel. *Std* refers to the protein standard *SeeBlue® Plus2 Pre-stained Protein Standard*, shown in kDa units. Samples A1, A2 and A3 are triplicates of the standard cultivation while A4, A5 and A6 are triplicates of the cultivation triggering increase of ncBCAA production.

A 3 point calibration curve was then elaborated for each gel by correlating added pre-proinsulin mass (1 μ g, 0.5 μ g and 0.25 μ g pure PPI) with its corresponding band intensity in the SDS-PAGE gel (Figure 24).

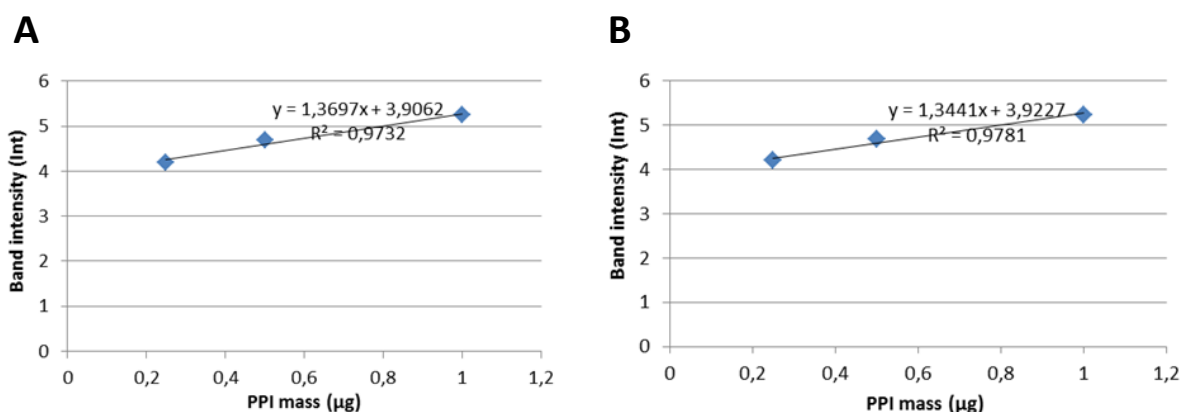


Figure 24. Calibration curve correlating added pre-proinsulin mass and its corresponding band intensity in the SDS-PAGE gel containing samples A1, A2, A3, A4, A5 (left) and A6 (right). Samples A1, A2 and A3 are triplicates of the standard cultivation while A4, A5 and A6 are triplicates of the cultivation triggering increase of ncBCAA production.

Afterwards, the intensity of the band corresponding to the expressed recombinant mini-proinsulin in a given sample was interpolated into the previously generated calibration curves and the corresponding PPI mass was then estimated. PPI concentrations resulting from the densitometry analysis are shown in Figure 25. The strategy based on pyruvate pulsing and O_2 limitation did not have a negative impact on recombinant protein production, as seen by the average values.

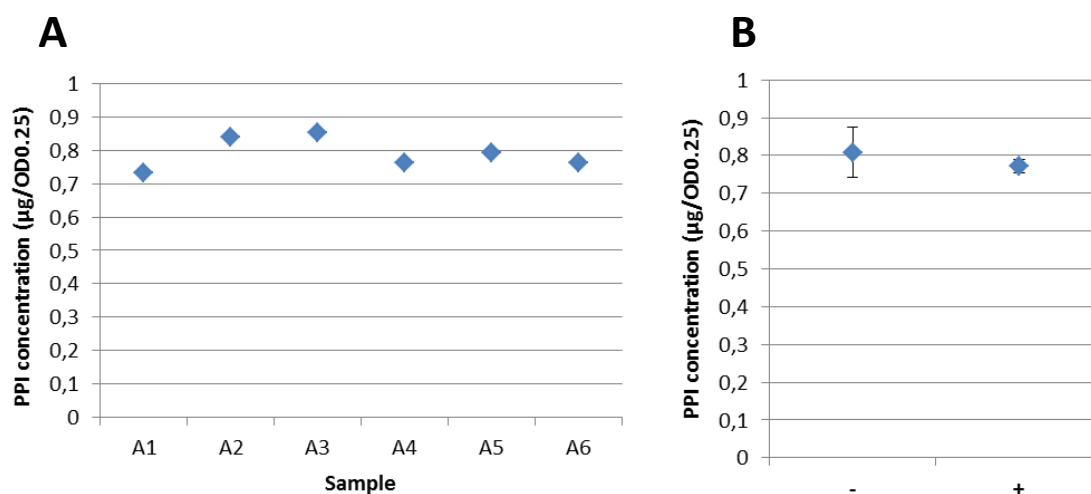


Figure 25. Estimated concentrations of recombinant mini-proinsulin 3h after induction of *E. coli* K-12 BW25113 pSW3_ *lact*⁺ cultivations from mini-reactor wells A1-A6 (A) and its average (B). Samples A1, A2 and A3 are triplicates of the standard cultivation (-) while A4, A5 and A6 are triplicates of the cultivation triggering increase of ncBCAA production (+).

5.7.1.3 Mini-proinsulin analysis by SDS-PAGE from inclusion body fractions

Culture broths from mini-reactor wells A1 to A6 were treated as previously described in section 4.8.1 in order to isolate inclusion bodies containing recombinant mini-proinsulin. In order to evaluate efficiency of the IB isolation process, samples after inclusion body isolation were analyzed by SDS-

PAGE as described in step 12 in section 4.8.1. Scanned 2D SDS-PAGE gels are shown in Figure 26A, where each gel lane corresponds to a cell suspension of $OD_{600nm}=0.5$. As a control, the total cell extract of an $OD_{600nm}=0.5$ cell suspension from mini-reactor A1 was employed.

By comparing cell extract and inclusion body samples it could be concluded that through the inclusion body isolation process there was no significant loss of recombinant mini-proinsulin, since the corresponding PPI band remained quite similar, and that recombinant mini-proinsulin in the inclusion body fraction was quite pure. Additionally, a 3-dimensional (3D) image showing volumetric intensities of bands present in the gel was also taken in order to clearly see that the recombinant mini-proinsulin was the predominant protein in the inclusion body fraction (Figure 26B).

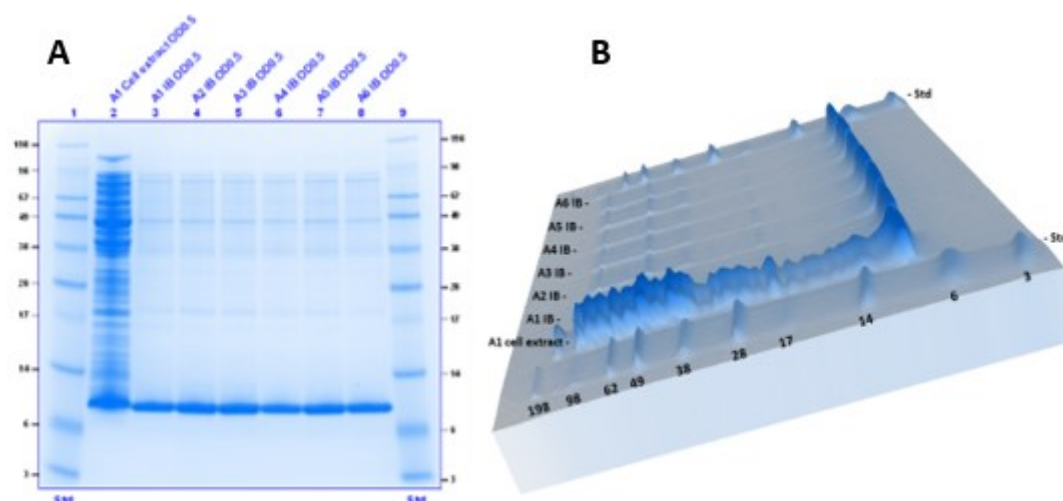


Figure 26. SDS-PAGE gel of inclusion body fraction from samples A1 to A6 in 2D (A) and 3D (B). Each gel lane corresponds to a cell suspension of $OD_{600nm}=0.5$. In order to determine efficiency of the inclusion body isolation process a control consisting on total cell extract of an $OD_{600nm}=0.5$ cell suspension from mini-reactor A1 was included. *Std* refers to the protein standard *SeeBlue® Plus2 Pre-stained Protein Standard*, shown in kDa units. Samples A1, A2 and A3 are triplicates of the standard cultivation while A4, A5 and A6 are triplicates of the cultivation triggering increase of ncBCAA production.

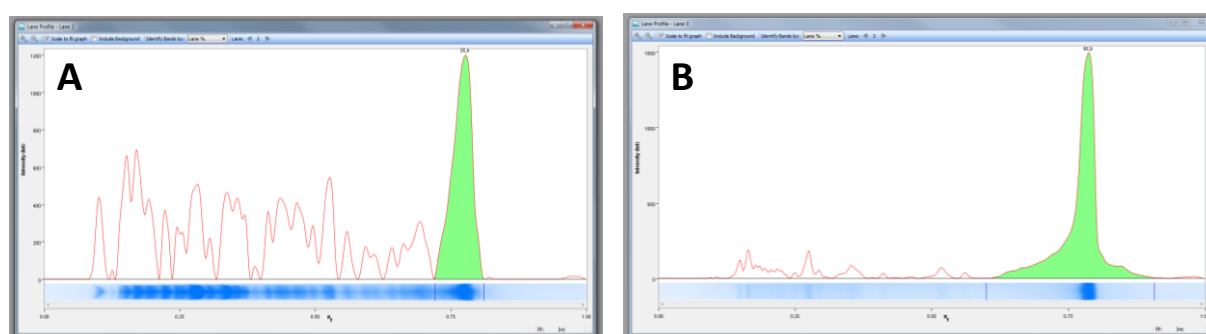


Figure 27. Chart showing band intensity (y-axis) over lane position (x-axis) for SDS-PAGE lanes “A1 cell extract” (A) and “A1 IB” (B). Lane region containing expressed recombinant mini-proinsulin was selected (highlighted in green) and the percentage of lane intensity represented by the recombinant protein was estimated, obtaining 25.4 % for (A) and 80.9% for (B).

For lanes “A1 cell extract” and “A1 IB” a chart showing intensity over lane position was also generated in order to estimate the percentage of the lane intensity represented by the recombinant

mini-proinsulin (Figure 27). In the cell extract, mini-proinsulin represented about 25% of the total protein while, after inclusion body isolation, this percentage increased up to about 81%. After results shown in this section, it was concluded that efficiency of performed inclusion body isolation process was high, obtaining a quite pure recombinant protein for further amino acid analysis.

5.7.1.4 Analysis of ncBCAA

Intracellular soluble protein fraction and inclusion body isolation from total cell extracts was carried out according to described in section 4.8.1. In the current case, at step 1, OD_{600nm} at the end of the cultivation was determined and all samples were standardized so that OD_{600nm} was equal to 43.7 in 1 mL. The corresponding volume was calculated for each sample and cells were harvested from liquid culture by centrifugation at 10000 g for 10 min. Acid hydrolysis of intracellular soluble and inclusion body protein fractions was performed as described in section 4.8.2. In the current case, 250 µL of the intracellular soluble protein fraction were mixed with 750 µL 5M HCl. Isolated inclusion body pellets were resuspended with 200 µL H₂O. 100 µL of the resulting inclusion body suspension were mixed with 900 µL 5M HCl. Hydrolysed IB pellet samples were resuspended with 500 µL dH₂O while hydrolysed intracellular soluble protein fraction samples were resuspended with 1 mL dH₂O. The resulting hydrolyzed samples were treated according to described in section 4.8.3. In the current case, 200 µL of the resulting IB solution were mixed with 100 µL 200 µM ABA while 100 µL of the resulting solution containing intracellular soluble protein fraction were mixed with 100 µL 200 µM ABA. The resulting samples were used for amino acid analysis by GC-FID according to the procedure described in sections 4.8.4-4.8.5. Concentrations of ncBCAA in both inclusion body and intracellular protein soluble fractions under the 2 tested cultivation conditions are shown in Figure 28.

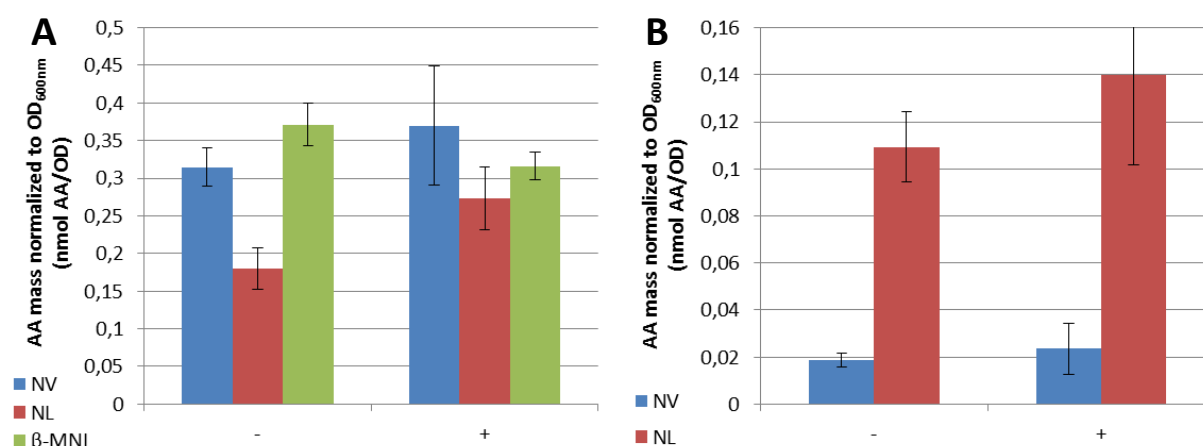


Figure 28. Molar concentrations of norvaline (blue bars), norleucine (red bars) and β-methylnorleucine (green bars) normalized to OD_{600nm} present in the intracellular soluble protein fraction (A) and inclusion body fraction (B) at 3 h after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation in a 10 mL mini-reactor under standard cultivation conditions (-) or under conditions triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+).

After implementing pyruvate pulses combined with O₂ limitation at mini-reactor level, norvaline increased a 17.5 %, norleucine increased a 51.7 % and β-methylnorleucine decreased a 14.8 % in the intracellular soluble protein fraction while norvaline increased a 25.6 % and norleucine increased a

28.1 % in the inclusion body fraction when normalizing amino acid concentration to OD_{600nm} (Figure 28). After implementing pyruvate pulses combined with O₂ limitation at mini-reactor level, norvaline increased a 73.7 % while norleucine increased a 32.9 % in the inclusion body fraction when normalizing amino acid concentration to PPI mass (Figure S71). Similar results were obtained when expressing data as ratio of the concentration of ncBCAA with respect to the canonical counterpart (Figure S72).

5.7.2 Experiment in a 15 L reactor

5.7.2.1 Cultivation mode

5.7.2.1.1 Standard cultivation

100 µL of a cryostock containing *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ were used to inoculate 500 mL of supplemented TUB mineral salt medium containing 5 g/L glucose and 100 µg/mL ampicillin in order to generate the pre-culture. The pre-culture was incubated at 37 °C and 220 rpm in an orbital shaker for 12 h, using an initial cold-start technique. OD_{600nm} at the end of the pre-culture was measured and a given volume was used to inoculate a 7 L starting volume reactor so that initial OD_{600nm} was 0.4. The reactor medium consisted of supplemented TUB mineral salt medium containing 5 g/L glucose, 2 mL antifoam (Antifoam 2014, Sigma) and 100 µg/mL ampicillin. Cultivation was carried out at 37 °C and the pH was maintained at 7 by automatic control with 25% NH₄OH. Airflow was set to 7 vvm and DO set-point to 20 %, maintaining the last by using a cascade control altering stirrer speed (initial stirrer speed was set to 800 rpm). Batch phase lasted effectively 4h, with an intermediate 13h cold phase at 15 °C. At the end of the batch phase, exponential feeding was started, according to following equation,

$$F(t) = \frac{q_s}{S} \cdot (X \cdot V) \cdot e^{\mu_{set} \cdot t}$$

where $F(t)$ represents the feed flow rate over time (L h⁻¹), q_s the set-point of the specific substrate uptake rate (0.514 gS gX⁻¹ h⁻¹), S the concentration of glucose in the feed solution (442 g/L), X the biomass concentration over time (g/L), V the volume of the reactor over time (L), μ_{set} the set-point of the specific cell growth rate (0.3 h⁻¹) and t the time during the fed-batch phase. The feed solution consisted of TUB mineral salt medium supplemented with 4 mL/L trace elements solution, 2 mL/L MgSO₄ solution (1.0 M), 100 µg/mL ampicillin and 442 g/L glucose.

Exponential fed-batch phase was carried out for 3 hours and afterwards expression of recombinant mini-proinsulin was induced by automatic addition of IPTG to a final concentration of 0.5 mM. Induction time was 30 minutes. During the induction phase no feed was added into the reactor. After induction, a constant feeding phase was started, so that the constant flow rate was equal to the last flow rate achieved in the exponential feeding phase. Constant feed fed-batch phase was active for 5-6 h. Glucose concentration in the reactor was determined by using a BioPAT® Trace online glucose analyser. A general overview of the cultivation is shown in Figure 29.

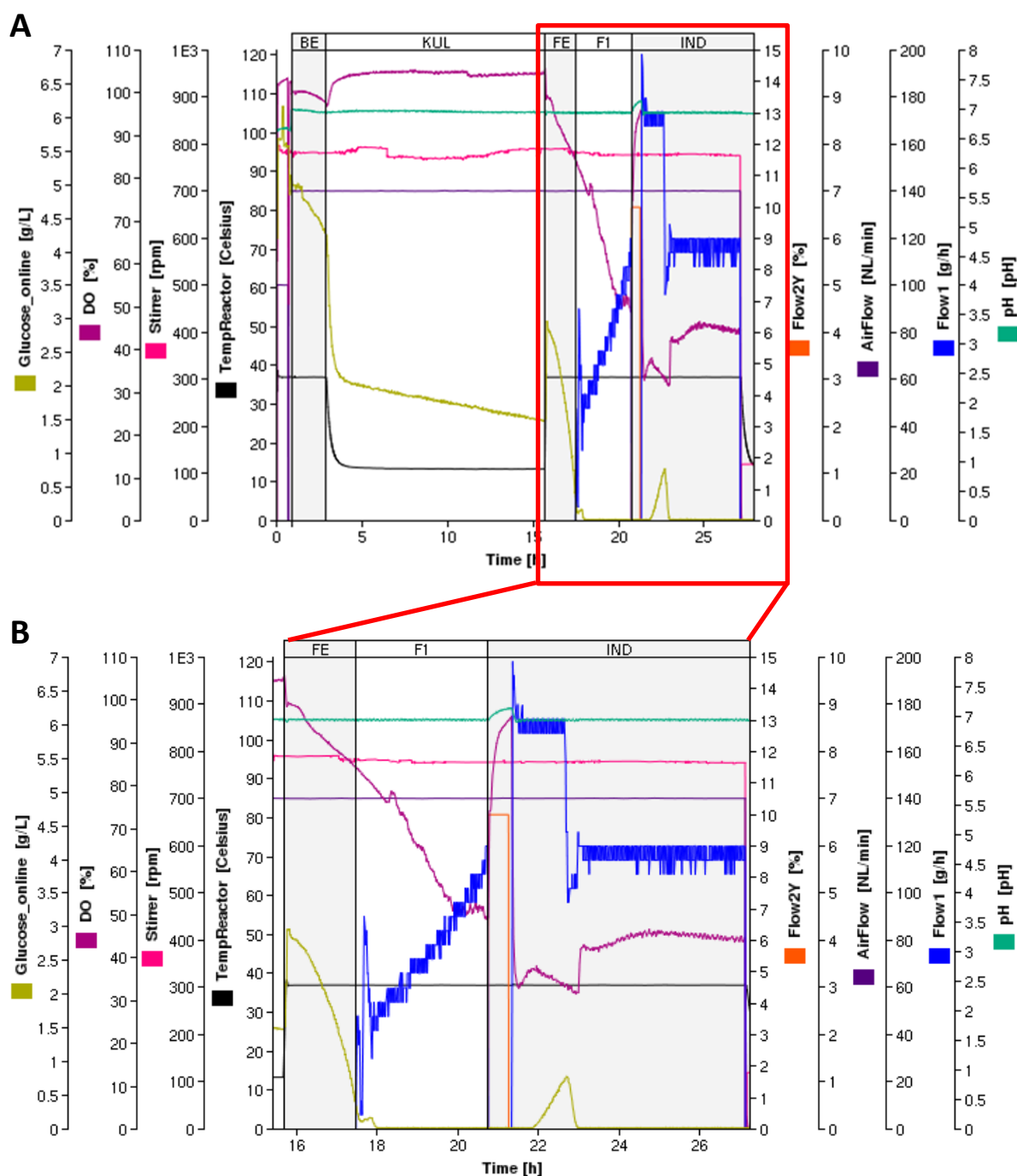


Figure 29. Overview of the standard cultivation of *E. coli* K-12 BW25113 pSW3_lacI* in a 15L reactor, during the whole cultivation process (A) and during fed-batch period (B). Different cultivation phases are shown in the diagram as BE (first 2h of batch phase), KUL (13h cold period at 15 °C), FE (remaining 2h of batch phase), F1 (3h exponential fed-batch phase) and IND (induction and linear fed-batch phase). IPTG induction was performed during 30 minutes (20.75 to 21.25 h cultivation time). Present in the diagram axes, Flow1 corresponds to the flow rate (g/h) of the pump transporting the feed solution into the reactor while Flow2Y corresponds to one tenth of the flow rate (g/h) of the pump transporting the IPTG solution used for induction. During the linear feeding phase the constant flow rate was supposed to be equal to the last flow rate achieved in the exponential feeding phase; however, during the beginning of the linear fed-batch phase that flow rate was higher due to a programming error. That explains the sudden accumulation of glucose (up to 0.8 g/L) between 22 and 23h cultivation time. Nevertheless, error was corrected by setting the flow rate to the proper value and glucose limitation was then again rapidly restored.

5.7.2.1.2 Cultivation triggering increase of ncBCAA production

Cultivation was performed as described for the standard cultivation in section 5.7.2.1.1. However, in the current cultivation the feed solution contained 454 g/L glucose instead of 442 g/L. Moreover, after the exponential fed-batch phase, 1 g/L pyruvate pulse was automatically added into the reactor. Pyruvate solution was constantly pumped for 5 minutes. During that time period no feed was added, airflow rate was temporary set to 0 and DO cascade control was disconnected. After the first pyruvate pulse, expression of recombinant mini-proinsulin was induced by automatic addition of IPTG to a final concentration of 0.5 mM. Induction time was 30 minutes. During the induction phase no feed was added into the reactor and airflow and DO cascade control were re-established. After induction, sequential 1 g/L pyruvate pulses were performed each 30 min as described above for a total of 4 pulses. Between pulses, constant feeding phase was activated, so that the constant flow rate was equal to the last flow rate achieved in the exponential feeding phase, and airflow and DO cascade control were re-established. Constant feed fed-batch phase was active for 5-6 h. A general overview of the cultivation is shown in Figure 30.

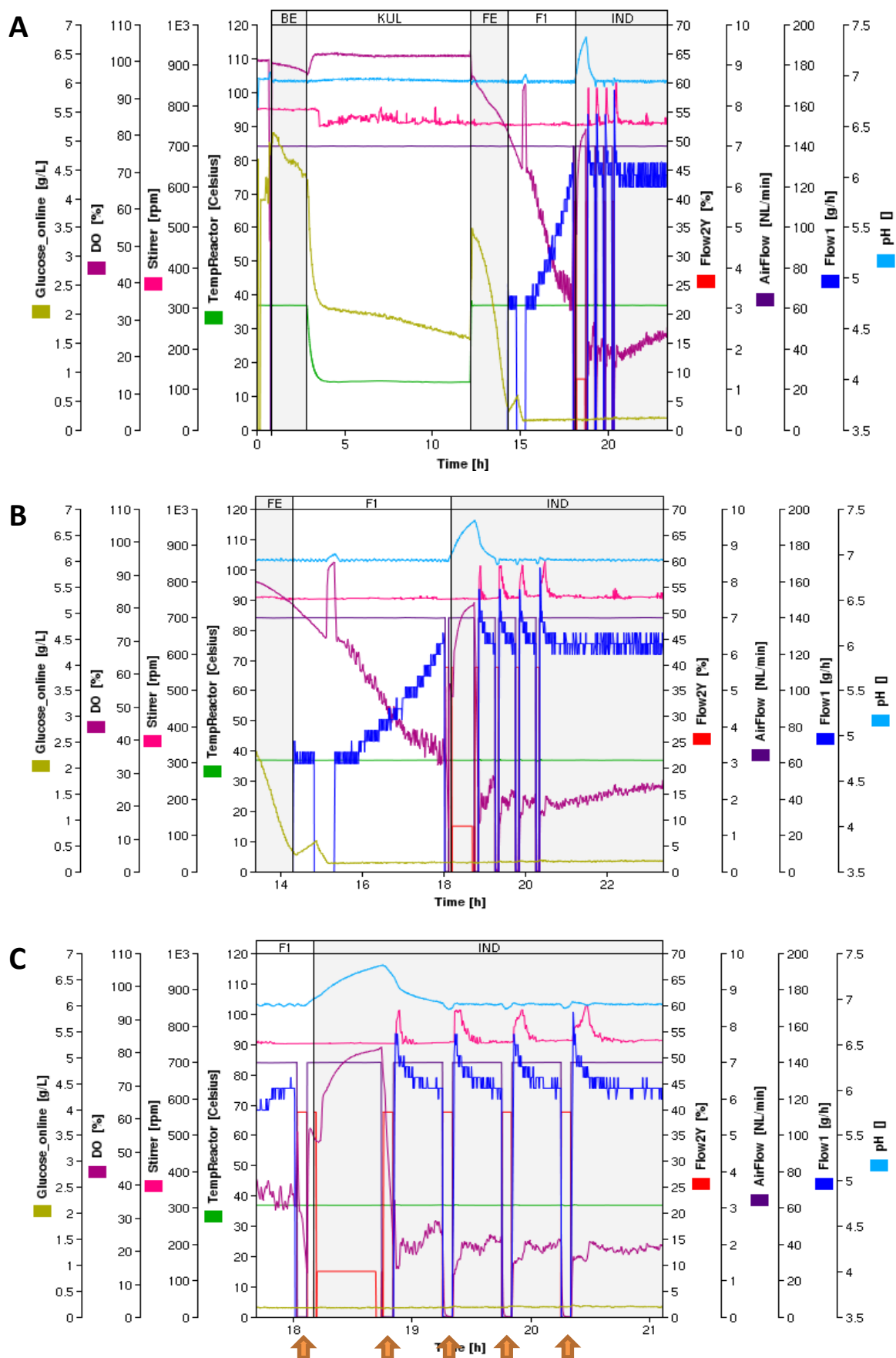


Figure 30. Overview of the cultivation of *E. coli* K-12 BW25113 pSW3_lacI* in a 15L reactor under conditions triggering increase of ncBCAA production, during the whole cultivation process (A), during fed-batch period (B)

and during pyruvate pulsing (C). Different cultivation phases are shown in the diagram as *BE* (first 2h of batch phase), *KUL* (10h cold period at 15 °C), *FE* (remaining 2h of batch phase), *F1* (3h exponential fed-batch phase) and *IND* (induction, linear fed-batch phase and pyruvate pulsing). IPTG induction was performed for 30 minutes (18.25 to 18.75 h cultivation time). Pyruvate pulses are indicated by orange arrows (C). Present in the diagram axes, *Flow1* corresponds to the flow rate (g/h) of the pump transporting the feed solution into the reactor while *Flow2Y* corresponds to one tenth of the flow rate (g/h) of the pump transporting either the IPTG solution used for induction or the pyruvate solution employed for pulsing. Exponential fed-batch was, due to a programming error, started shortly before glucose was completely consumed in the batch phase. Hence a small glucose accumulation was reported at around 15h. Thus, exponential feeding was shortly shut down until glucose was completely depleted and then, activated again.

5.7.2.2 Mini-proinsulin analysis by HPLC

Concentration of recombinant mini-proinsulin from hourly samples was analyzed according to an *in house* HPLC method. 1 mL of culture broth was directly used for analysis. PPI concentrations resulting from the *in house* HPLC analytical method are shown in Figure 31. PPI concentrations determined under both tested cultivation conditions were really similar. In addition, cell growth behavior (represented by OD_{600nm} and CDW) was also comparable for both tested cultivation conditions (Figure S73), suggesting that specific production of recombinant mini-proinsulin remained the same, independently of the cultivation strategy employed.

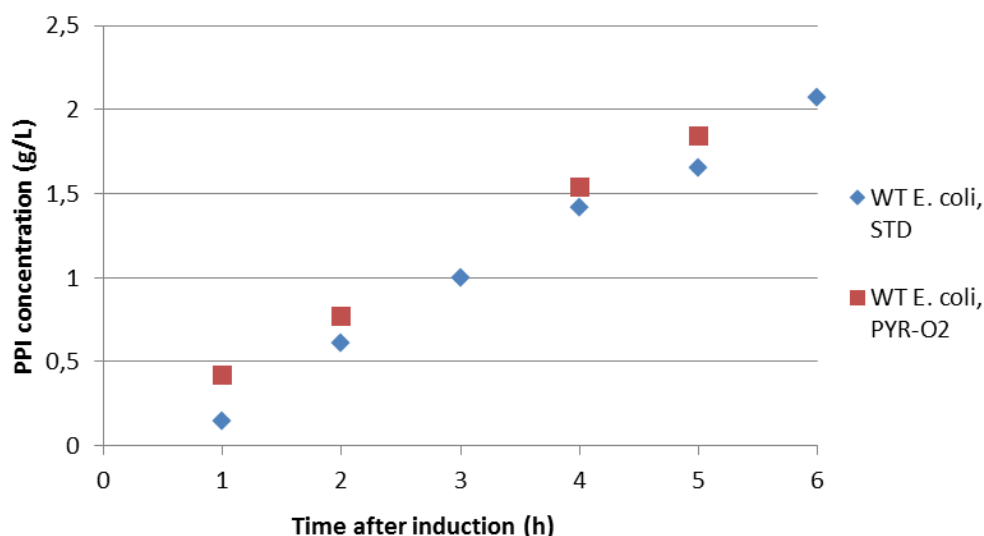


Figure 31. Estimated concentrations of recombinant mini-proinsulin over time after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation (WT *E. coli*) in a 15L reactor under standard conditions (STD) and under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2).

5.7.2.3 Analysis of ncBCAA

Intracellular soluble protein fraction and inclusion body isolation from total cell extracts was carried out according to described in section 4.8.1. In the current case, at step 1, the OD_{600nm} of the culture at a given time point was determined and all samples were standardized so that OD_{600nm} was equal to 50 in 1 mL. The corresponding volume was calculated for each sample and cells were harvested from liquid culture by centrifugation at 10000 g for 10 min. Acid hydrolysis of intracellular soluble protein and inclusion body fractions was performed as described in section 4.8.2. In the current case, 250 µL of the intracellular soluble protein fraction were mixed with 750 µL 5M HCl. Isolated inclusion body

pellets were resuspended with 200 μL H_2O . 100 μL of the resulting inclusion body suspension were mixed with 900 μL 5M HCl. After acid hydrolysis and speed-back, hydrolysed IB pellet samples were resuspended with 250 μL dH_2O while hydrolysed intracellular soluble protein fraction samples were resuspended with 1 mL dH_2O . The resulting hydrolyzed samples were treated according to described in section 4.8.3. In the current case, 250 μL of the resulting IB solution (whole solution) were mixed with 100 μL 200 μM ABA while 100 μL of the resulting solution containing intracellular soluble protein fraction were mixed with 100 μL 200 μM ABA. The resulting samples were used for amino acid analysis by GC-FID according to the procedure described in sections 4.8.4-4.8.5. Concentrations of ncBCAA in the intracellular soluble protein and inclusion body fraction over cultivation time are shown in Figure 32 and Figure 33, respectively.

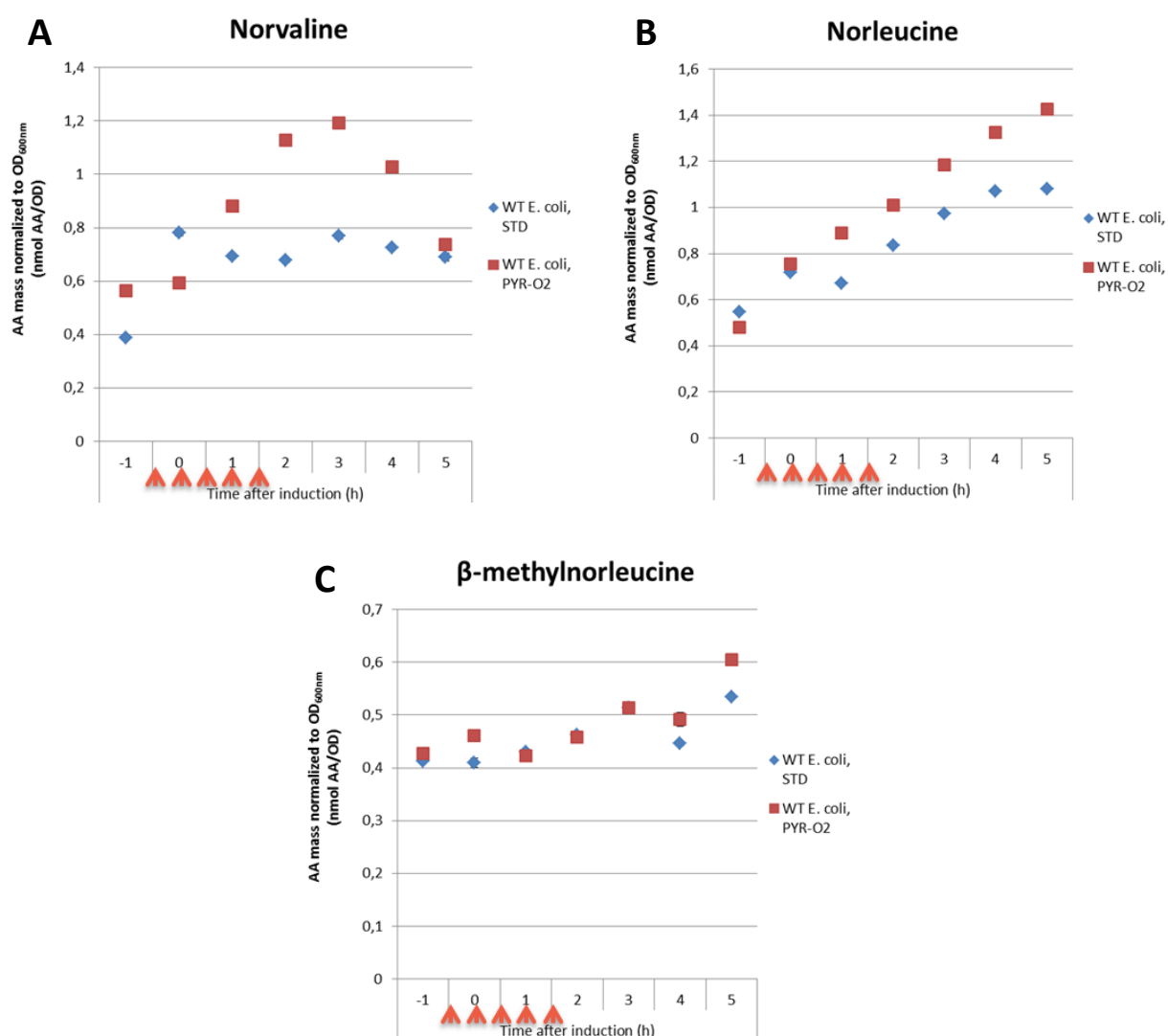


Figure 32. Molar concentrations of norvaline (A), norleucine (B) and β -methylnorleucine (C) normalized to $\text{OD}_{600\text{nm}}$ present in the intracellular soluble protein fraction calculated over time after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation (WT *E. coli*) in a 15L reactor under standard conditions (STD) and under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2). Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O_2 limitation was applied.

For both tested cultivations, a progressive accumulation of norleucine and β -methylnorleucine was reported in the intracellular soluble protein fraction over time after induction, being it more significant for norleucine. However, for standard cultivation “WT *E.coli*, STD”, after an initial increase reported at 0 h induction time, norvaline concentration remained quite invariable over time (Figure 32).

Cultivation subjected to pyruvate pulses and O₂ limitation (“WT *E.coli*, PYR-O2”) translated in an increase of ncBCAA with respect to cultivation “WT *E.coli*, STD”. Depending on the analyzed ncBCAA, such increase was reported at different time points after first pyruvate pulse was applied. Norvaline and norleucine concentrations increased in cultivation “WT *E.coli*, PYR-O2” with respect to “WT *E.coli*, STD” 1.5 h after first pyruvate pulse was applied (1 h after induction) and such increase persisted for the remaining cultivation time. However, increase of β -methylnorleucine concentration was reported from 4.5 h after first pyruvate pulse (4 h after induction) on. This suggests that pyruvate is not immediately metabolized after its addition into the reactor and that the metabolic effect triggered by pyruvate and O₂ limitation has a limited influence on β -methylnorleucine synthesis, in contrast to norleucine and norvaline (Figure 32).

Furthermore, norvaline concentration increased progressively in cultivation “WT *E.coli*, PYR-O2” until 3h after induction. However, from that time point on, norvaline concentration progressively dropped to finally reach values comparable to standard cultivation “WT *E.coli*, STD” 5 h after induction. Hence, the norvaline increase resulting upon applying pyruvate pulses and O₂ limitation disappeared 2h after last pyruvate pulse was applied. Interestingly, this observation was not shown for norleucine and β -methylnorleucine, which reported a progressive accumulation over cultivation time (Figure 32).

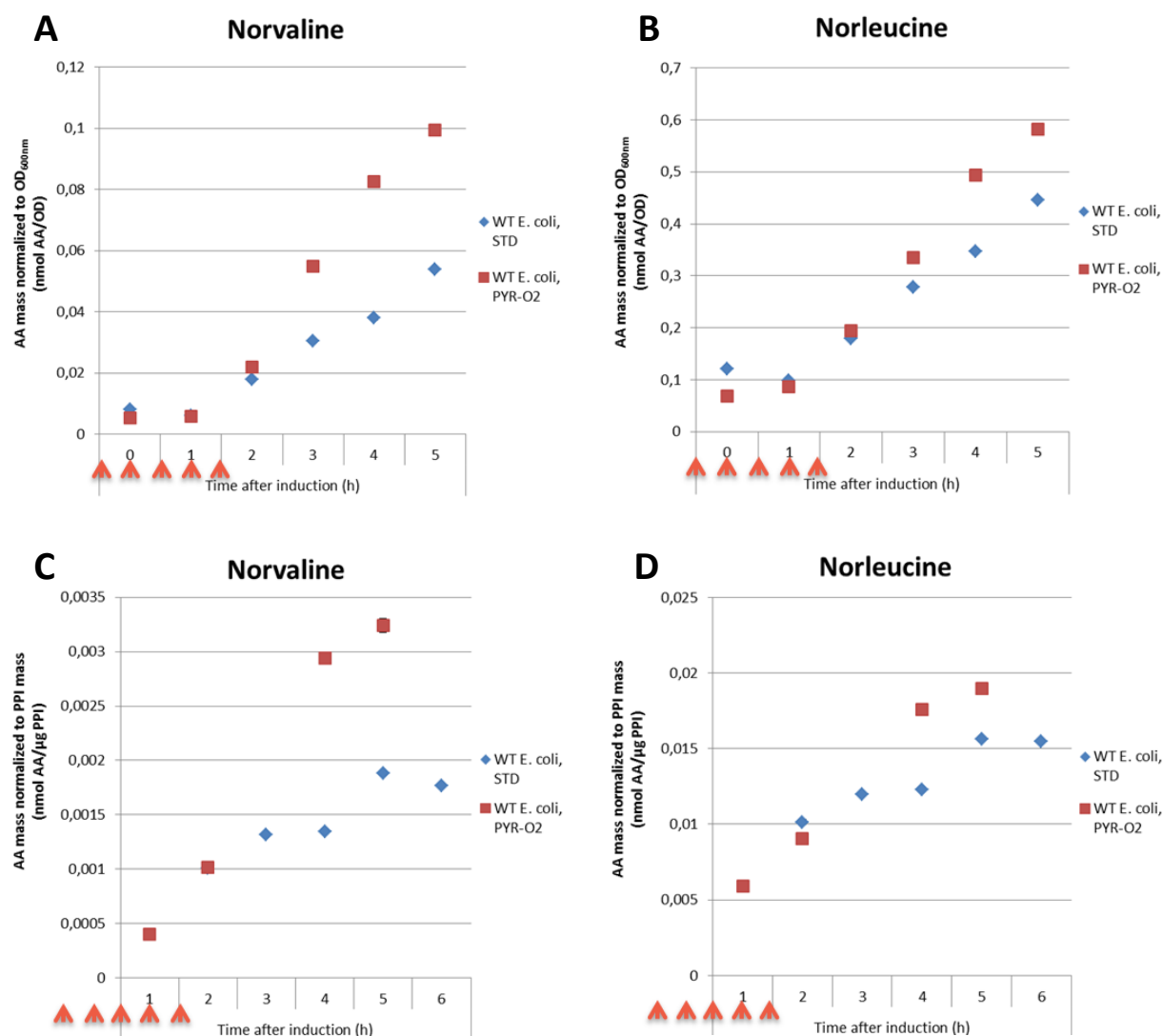


Figure 33. Molar concentrations of norvaline (A, C) and norleucine (B, D) normalized to OD_{600nm} (A, B) or to PPI mass (C, D) present in the inclusion body fraction calculated over time after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation (WT *E. coli*) in a 15L reactor under standard conditions (STD) and under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2). Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O_2 limitation was applied.

For both tested cultivations, a progressive accumulation of norvaline and norleucine was reported in the inclusion body fraction over time after induction. Until 2 h after induction, norvaline and norleucine concentrations remained similar in both tested cultivations. However, from that time point on, norvaline and norleucine concentrations analyzed in cultivation “WT *E.coli*, PYR-O2” remained higher than in “WT *E.coli*, STD” (Figure 33). Hence, the increase of norvaline and norleucine concentration resulting upon applying pyruvate pulses and O_2 limitation was first reported 3.5 h after first pyruvate pulse was applied (3 h after induction), which is 2 h later than reported in the intracellular soluble protein fraction. This suggests that metabolic effects triggered by pyruvate pulsing combined with O_2 limitation occur first in the cytosol (intracellular soluble protein fraction), where the ncBCAA are synthesized. Afterwards, translation machinery would mis-incorporate those ncBCAA present in the cytosol into the nascent recombinant proteins (inclusion body fraction). β -

methylnorleucine could not be properly detected in the inclusion body fraction since at the expected retention time in the chromatogram no peak was found (Figure 33).

5.7.2.4 Acetate and formate analysis

Acetate and formate concentrations from hourly samples were analysed offline by *in house* enzymatic assays available at Sanofi-Aventis Deutschland GmbH. Sample preparation for acetate and formate analysis was as follows: 10 mL of culture broth were inactivated with Bardac, centrifuged at 4000 g for 10 min and resultant supernatant was decanted and filtered through a 0.45 μ m filter. The filtrate was then used for analysis. Estimated concentrations of acetate and formate in culture broth supernatants over cultivation time are shown in Figure 34.

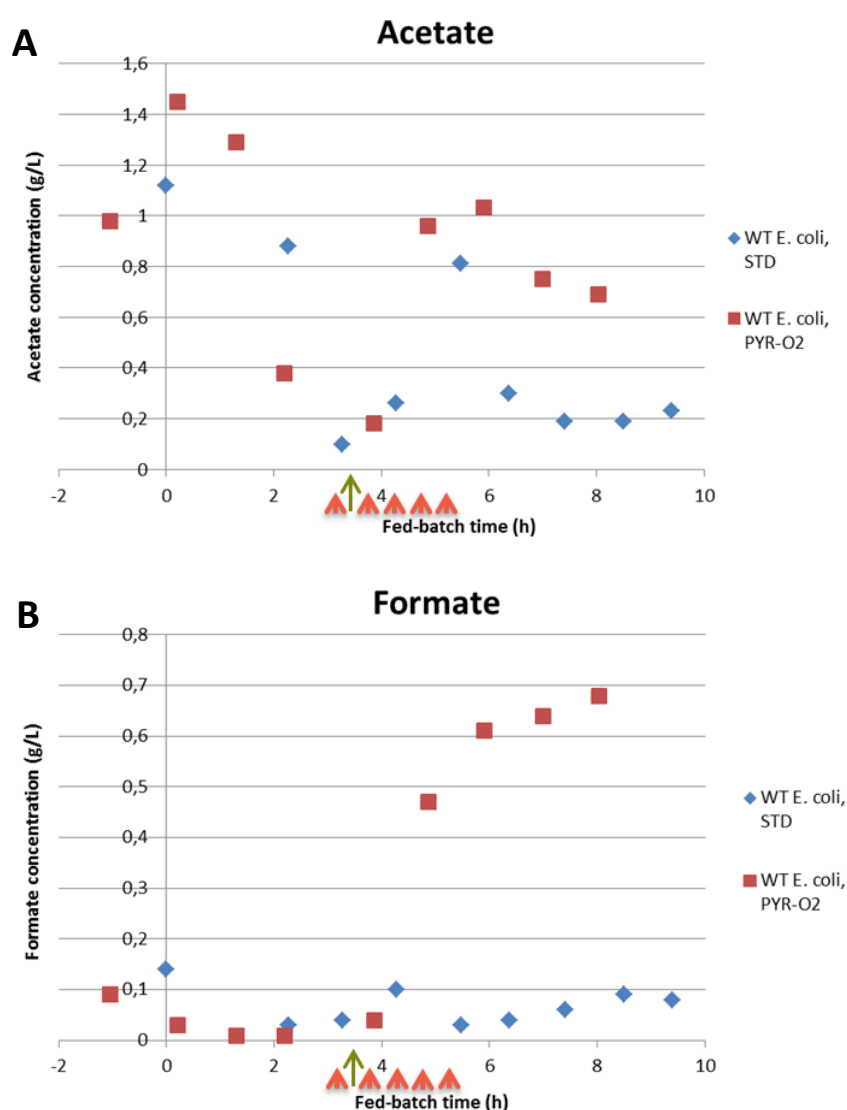


Figure 34. Concentration of acetate (A) and formate (B) present in culture broth supernatants calculated over fed-batch time of cultivation *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ (WT *E. coli*) in a 15L reactor under standard cultivation conditions (STD) and under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2). Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O₂ limitation was applied. Green arrow points out time point where IPTG induction was carried out.

For both tested cultivations, acetate concentration reached a maximum after batch phase was finished (0 h fed-batch time). For standard cultivation “WT *E.coli*, STD”, acetate concentration strongly decreased during the first hours of the fed-batch phase, reaching the minimum at around 4 h fed-batch time, to suddenly increase again, reporting a maximum at 5.5 h fed-batch time. Afterwards, acetate decreased again and remained invariable at values around 0.2 g/L until the end of the fed-batch phase. That sudden and momentary accumulation of acetate during the fed-batch phase in cultivation “WT *E.coli*, STD” corresponds with the sudden glucose accumulation reported at the beginning of the linear fed-batch phase due to a programming error (Figure 29). Until 4 h fed-batch time, acetate concentration in cultivation “WT *E.coli*, PYR-O2” remained quite similar to “WT *E.coli*, STD”. However, from that time point on, acetate concentration was always higher for cultivation “WT *E.coli*, PYR-O2”, as a consequence of the metabolic alteration triggered by pyruvate pulsing combined with O₂ limitation.

For both cultivations tested, formate concentration reached a maximum after batch phase was finished (around 0 h fed-batch time). During the first hours of fed-batch phase in standard cultivation “WT *E.coli*, STD”, formate concentration decreased, but then a sudden increase was reported at 3.3 and 4.3 h fed-batch time. Afterwards, formate concentration decreased again, remaining below 0.1 g/L. Like stated for acetate, such momentary accumulation of formate in cultivation “WT *E.coli*, STD” is in accordance with the sudden glucose accumulation reported at the beginning of the linear fed-batch phase due to a programming error (Figure 29). Until 4 h fed-batch time, formate concentration in cultivation “WT *E.coli*, PYR-O2” remained quite similar to “WT *E.coli*, STD”. However, from that time point on, formate concentration was always higher for cultivation “WT *E.coli*, PYR-O2”, as a consequence of the metabolic effect triggered by pyruvate pulsing combined with O₂ limitation. Such effect seems to be stronger for formate than acetate since, after pyruvate pulsing, increase of formate concentration in cultivation “WT *E.coli*, PYR-O2” with respect to standard cultivation “WT *E.coli*, STD” is about 700 %, while for acetate is about 250 %. It is also noteworthy to highlight that right after last pyruvate pulse combined with O₂ limitation was applied, acetate and formate concentrations suddenly stopped increasing, which might suggest that anaerobic metabolism is not active anymore.

Results derived from acetate and formate analysis are consistent with results shown in section 5.7.2.3 for norleucine and norvaline in the intracellular soluble protein fraction since effective increase of ncBCAA in cultivation “WT *E.coli*, PYR-O2” with respect to standard “WT *E.coli*, STD”, as reported with acetate and formate, took place 1.5 h after addition of the first pyruvate. Hence, acetate and formate analysis were proven to be a good additional marker of ncBCAA production in *E. coli* cultivations.

5.7.3 Comparison 10 mL mini-reactor and 15 L reactor

In order to elucidate if scale has an impact on ncBCAA production, a comparative study summarizing results obtained in both scales (10 mL mini-reactor and 15 L reactor) was elaborated. Results obtained from samples taken 3 h after induction were compared, since that time point was the only selected for ncBCAA analysis at mini-reactor scale. In both intracellular soluble protein fraction and inclusion body fraction, ncBCAA concentrations obtained at mini-reactor and 15 L reactor scale under standard cultivation conditions or under conditions triggering ncBCAA production (i.e. pyruvate pulsing combined with O₂ limitation) were compared (Figure 35A and Figure 36A). In addition, for

both tested scales, percentage of variation of ncBCAA concentration in cultivation triggering ncBCAA production with respect to standard cultivation was determined (Figure 35B and Figure 36B).

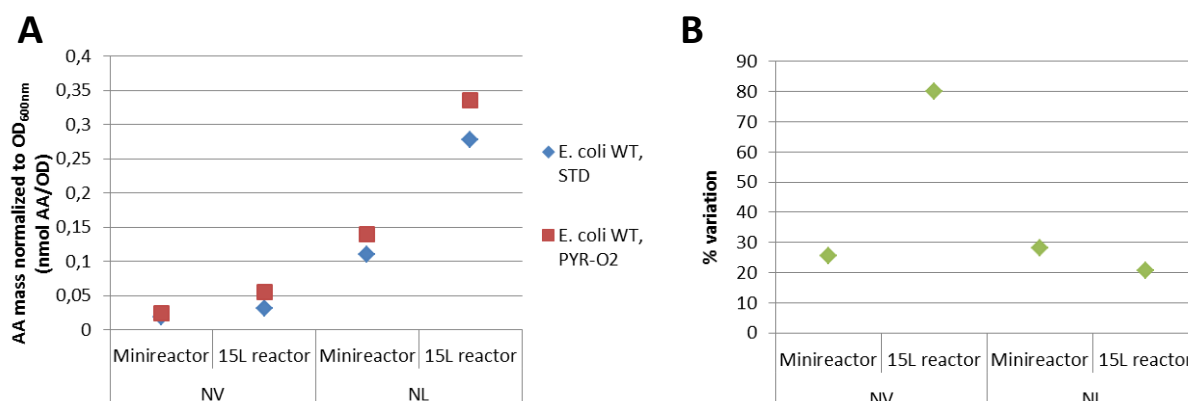


Figure 35. (A) Comparison of ncBCAA concentration calculated 3h after induction in the inclusion body fraction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ (WT *E. coli*) cultivation in 10 mL mini-reactor- and 15L reactor scale under standard cultivation conditions (STD) or under conditions triggering ncBCAA production, i.e. pyruvate pulsing combined with O₂ limitation (PYR-O2). (B) Percentage of variation of ncBCAA concentration reported in the cultivation triggering ncBCAA production with respect to standard cultivation.

In the inclusion body fraction, for both norvaline and norleucine, ncBCAA concentration values were higher at 15 L reactor than at mini-reactor scale (Figure 35A). That might be explained since at bigger scales inhomogeneities can generate different O₂ and glucose gradients, hence triggering formation of ncBCAA, even under standard cultivation conditions. In addition, effect of cultivation conditions triggering ncBCAA production based on pyruvate pulsing combined with O₂ limitation was reported to be different depending on the tested scale, especially for norvaline (Figure 35B): concentration of norvaline reported an increase of 26 % at mini-reactor scale while an 80 % boost was shown at 15 L reactor scale. Norleucine concentration increased 28 % at mini-reactor scale but 21 % at 15 L reactor scale.

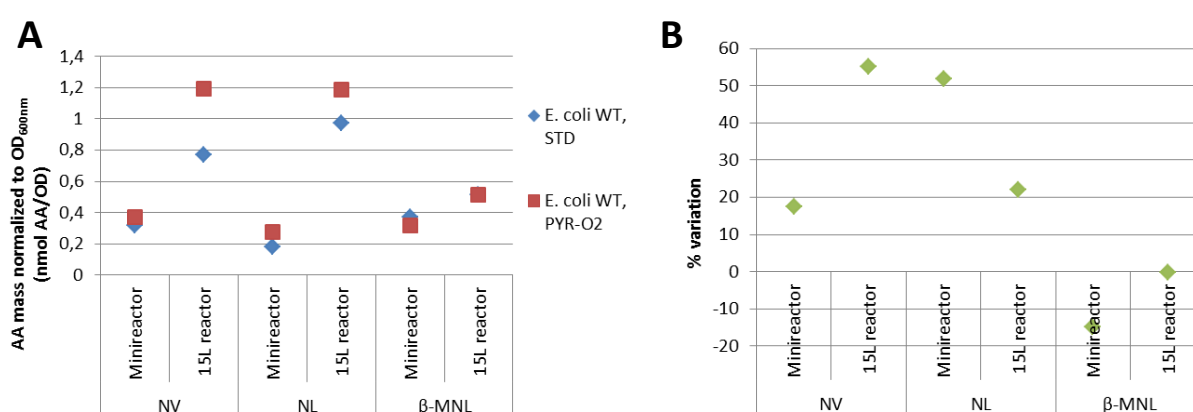


Figure 36. (A) Comparison of ncBCAA concentration determined 3h after induction in the intracellular soluble protein fraction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ (WT *E. coli*) cultivation in 10mL mini-reactor- and 15L reactor scale under standard cultivation conditions (STD) or under conditions triggering ncBCAA production, i.e. pyruvate pulsing combined with O₂ limitation (PYR-O2). (B) Percentage of variation of ncBCAA concentration reported in the cultivation triggering ncBCAA production with respect to standard cultivation.

As reported above, in the intracellular soluble protein fraction, for both norvaline and norleucine, concentration values were higher at 15 L reactor than at mini-reactor scale as well (Figure 36A). Also, effect of cultivation conditions based on pyruvate pulsing combined with O₂ limitation was reported to dramatically vary depending on the tested scale (Figure 36B): concentration of norvaline reported an increase of 17 % at mini-reactor scale while a 51 % boost was shown at 15 L reactor scale. Norleucine concentration increased 52 % at mini-reactor scale but only 22 % at 15 L reactor scale. However, for β -methylnorleucine effect was not so significantly different between scales when compared with the other ncBCAA: β -methylnorleucine concentration decreased 14 % at mini-reactor scale and reported no variation at 15 L reactor scale.

5.8 Screening of *geneX*-tunable *E. coli* strains a mini-reactor system

The aim of this experiment was to regulate the expression of strains *E. coli* K-12 BW25113 Δ *geneX* expressing pSW3_ *lacI*⁺ and pACG_araBAD_ *geneX* (simplified, *geneX*-tunable *E. coli* strains) by using different concentrations of L-arabinose in a 10 mL mini-reactor. Tunable *E. coli* strains were grown in Enbase-based fed-batch modus while carrying out pyruvate pulsing followed by O₂ limitation, conditions which in section 5.7 demonstrated to trigger production of ncBCAA. Tunable *E. coli* strains were also tested under standard cultivation conditions. By comparing the ncBCAA production profile obtained for each of the tested tunable *E. coli* strains with the control non-engineered *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ strain, it could be elucidated if the corresponding tunable *E. coli* strain (and more specifically, which level of expression of the corresponding target gene) improves the product quality of the expressed recombinant mini-proinsulin.

5.8.1 Cultivation conditions

Pre-cultures were prepared as follows: 30 μ L of a cryostock containing *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ or one of the *geneX*-tunable *E. coli* strains were used to inoculate 30 mL of 1:3 supplemented TUB medium containing 5 g/L glucose, 0.1 M Na-Phosphate buffer and 100 μ g/mL ampicillin. For the tunable *E. coli* strains, medium also contained 25 μ g/mL chloramphenicol and the minimum L-arabinose concentration necessary to recover cell growth levels of the non-engineered *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ strain, previously tested in section 5.3.2 (summarized in

Table 25). Pre-cultures were incubated at 37 °C and 220 rpm, overnight.

Table 25. Minimal L-arabinose concentration necessary for the cultivation of each generated tunable *E. coli* strains allowing recovery of cell growth levels of the non-engineered *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ strain.

Strain	L-arabinose concentration (%)
<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺	0
<i>E. coli</i> K-12 BW25113 Δ <i>leuA</i> pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>leuA</i> (<i>leuA</i> -tunable <i>E. coli</i>)	0.1
<i>E. coli</i> K-12 BW25113 Δ <i>ilvC</i> pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>ilvC</i> (<i>ilvC</i> -tunable <i>E. coli</i>)	0.4
<i>E. coli</i> K-12 BW25113 Δ <i>thrA</i> pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>thrA</i> (<i>thrA</i> -tunable <i>E. coli</i>)	0.4

<i>E. coli</i> K-12 BW25113 $\Delta ilvA$ pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>ilvA</i> (<i>ilvA</i> -tunable <i>E. coli</i>)	0,05
<i>E. coli</i> K-12 BW25113 $\Delta ilvBN$ pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>ilvBN</i> (<i>ilvBN</i> -tunable <i>E. coli</i>)	0
<i>E. coli</i> K-12 BW25113 $\Delta ilvIH$ pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>ilvIH</i> (<i>ilvIH</i> -tunable <i>E. coli</i>)	0
<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ pACG_araBAD_ <i>ilvGM</i> (<i>ilvGM</i> -tunable <i>E. coli</i>)	0

Main culture was prepared differently, depending on the tested cultivation conditions:

- Standard cultivation conditions

OD_{600nm} at the end of the pre-cultivation was measured and a given volume was used to inoculate a 5 mL starting volume Pall Micro24 mini-reactor (Microreactor Technologies Inc.) so that initial OD_{600nm} was 0.4. The mini-reactor medium consisted of 1:3 supplemented TUB medium containing 4 g/L glucose, 100 µg/mL ampicillin, 25 µg/mL chloramphenicol (only for tunable *E. coli* strains) and 1 µL/mL Desmophen antifoam. Medium was also supplemented with different concentrations of L-arabinose. Cultivation was carried out at 37 °C and the pH was maintained at 7 by automatic control with NH₄OH and CO₂. Stirrer speed was set to 800 rpm and DO set-point to 25 %, maintaining the last by automatically increasing the oxygen flow into the mini-reactor. Batch phase lasted around 4h. After batch phase was finished, 1 mL 400 g/L EnPump 200 solution and 50 µL 3000 U/L amylase solution were manually added into the mini-reactor, hence starting the fed-batch phase. 30 min after beginning of the fed batch phase, expression of recombinant mini-proinsulin was induced by manual addition of an IPTG pulse to a final concentration of 0.5 mM. Fed-batch phase was active for 3.5 h.

- Cultivation conditions triggering ncBCAA formation

Cultivation was performed as described for the standard cultivation. However, immediately after beginning of the fed-batch phase, a 0.833 g/L pyruvate pulse was manually added into the reactor. During the following 5 min after pyruvate addition, DO set-point was set to 0, so that no oxygen was supplied into the mini-reactor during that period, hence ensuring oxygen limitation. 30 min after the first pyruvate pulse, expression of recombinant mini-proinsulin was induced by manual addition of an IPTG pulse to a final concentration of 0.5 mM. After induction, sequential 0.833 g/L pyruvate pulses were manually performed each 30 min as described above for a total of 5 pulses. Between pulses, DO set-point was re-established to 25 %. Fed-batch phase was active for 3.5 h.

Cultivations were performed in 2 mini-reactor plates, comprising a total of 48 wells. Table 26 and Table 27 show, for each tested cultivation, the well position, the tested strain, the L-arabinose concentration employed as well as the cultivation conditions applied.

Table 26. Overview of the different cultivation conditions tested in each well of the first mini-reactor plate with strains *E. coli* K-12 BW25113 pSW3_ *lacI*⁺, *leuA*-tunable *E. coli*, *ilvC*-tunable *E. coli* and *thrA*-tunable *E. coli*.

	1	2	3	4	5	6
A	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Standard	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Standard	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Standard	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Pyruvate + O ₂ lim.	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Pyruvate + O ₂ lim.	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Pyruvate + O ₂ lim.
B	<i>leuA</i> -tunable <i>E. coli</i> 0.1 % L-ara Standard	<i>leuA</i> -tunable <i>E. coli</i> 0.2 % L-ara Standard	<i>leuA</i> -tunable <i>E. coli</i> 0.4 % L-ara Standard	<i>leuA</i> -tunable <i>E. coli</i> 0.1 % L-ara Pyruvate + O ₂ lim.	<i>leuA</i> -tunable <i>E. coli</i> 0.2 % L-ara Pyruvate + O ₂ lim.	<i>leuA</i> -tunable <i>E. coli</i> 0.4 % L-ara Pyruvate + O ₂ lim.
C	<i>ilvC</i> -tunable <i>E. coli</i> 0.4 % L-ara Standard	<i>ilvC</i> -tunable <i>E. coli</i> 0.8 % L-ara Standard	<i>ilvC</i> -tunable <i>E. coli</i> 1.6 % L-ara Standard	<i>ilvC</i> -tunable <i>E. coli</i> 0.4 % L-ara Pyruvate + O ₂ lim.	<i>ilvC</i> -tunable <i>E. coli</i> 0.8 % L-ara Pyruvate + O ₂ lim.	<i>ilvC</i> -tunable <i>E. coli</i> 1.6 % L-ara Pyruvate + O ₂ lim.
D	<i>thrA</i> -tunable <i>E. coli</i> 0.4 % L-ara Standard	<i>thrA</i> -tunable <i>E. coli</i> 0.8 % L-ara Standard	<i>thrA</i> -tunable <i>E. coli</i> 1.6 % L-ara Standard	<i>thrA</i> -tunable <i>E. coli</i> 0.4 % L-ara Pyruvate + O ₂ lim.	<i>thrA</i> -tunable <i>E. coli</i> 0.8 % L-ara Pyruvate + O ₂ lim.	<i>thrA</i> -tunable <i>E. coli</i> 1.6 % L-ara Pyruvate + O ₂ lim.

Table 27. Overview of the different cultivations tested in each well of the second mini-reactor plate with strains *E. coli* BW25113 pSW3_ *lacI*⁺, *ilvH*-tunable *E. coli*, *ilvA*-tunable *E. coli*, *ilvBN*-tunable *E. coli* and *ilvGM*-tunable *E. coli*.

	1	2	3	4	5	6
A	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Standard	<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺ 0 % L-ara Pyruvate + O ₂ lim.	<i>ilvH</i> -tunable <i>E. coli</i> 0.05 % L-ara Standard	<i>ilvH</i> -tunable <i>E. coli</i> 0.8 % L-ara Standard	<i>ilvH</i> -tunable <i>E. coli</i> 0.05 % L-ara Pyruvate + O ₂ lim	<i>ilvH</i> -tunable <i>E. coli</i> 0.8 % L-ara Pyruvate + O ₂ lim
B	<i>ilvA</i> -tunable <i>E. coli</i> 0.05 % L-ara Standard	<i>ilvA</i> -tunable <i>E. coli</i> 0.2 % L-ara Standard	<i>ilvA</i> -tunable <i>E. coli</i> 0.8 % L-ara Standard	<i>ilvA</i> -tunable <i>E. coli</i> 0.05 % L-ara Pyruvate + O ₂ lim	<i>ilvA</i> -tunable <i>E. coli</i> 0.2 % L-ara Pyruvate + O ₂ lim	<i>ilvA</i> -tunable <i>E. coli</i> 0.8 % L-ara Pyruvate + O ₂ lim
C	<i>ilvBN</i> -tunable <i>E. coli</i> 0.05 % L-ara Standard	<i>ilvBN</i> -tunable <i>E. coli</i> 0.2 % L-ara Standard	<i>ilvBN</i> -tunable <i>E. coli</i> 0.8 % L-ara Standard	<i>ilvBN</i> -tunable <i>E. coli</i> 0.05 % L-ara Pyruvate + O ₂ lim	<i>ilvBN</i> -tunable <i>E. coli</i> 0.2 % L-ara Pyruvate + O ₂ lim	<i>ilvBN</i> -tunable <i>E. coli</i> 0.8 % L-ara Pyruvate + O ₂ lim
D	<i>ilvGM</i> -tunable <i>E. coli</i> 0.05 % L-ara Standard	<i>ilvGM</i> -tunable <i>E. coli</i> 0.2 % L-ara Standard	<i>ilvGM</i> -tunable <i>E. coli</i> 0.8 % L-ara Standard	<i>ilvGM</i> -tunable <i>E. coli</i> 0.05 % L-ara Pyruvate + O ₂ lim	<i>ilvGM</i> -tunable <i>E. coli</i> 0.2 % L-ara Pyruvate + O ₂ lim	<i>ilvGM</i> -tunable <i>E. coli</i> 0.8 % L-ara Pyruvate + O ₂ lim

5.8.2 Mini-proinsulin analysis by SDS-PAGE

Determination of recombinant mini-proinsulin concentration after induction for each strain under each tested cultivation condition was carried out by densitometry analysis by SDS-PAGE exactly as described in section 5.7.1.2 (protein gels not shown). Results are shown in Table S8 and Table S9, corresponding to mini-reactor plates presented at Table 26 and Table 27, respectively.

5.8.3 ncBCAA analysis

Samples for amino acid analysis by GC-FID were prepared exactly as aforementioned in section 5.7.2.3. Concentrations of ncBCAA in both intracellular soluble protein fraction and inclusion body fraction for each *E. coli* K-12 BW25113 $\Delta lgeneX$ pSW3_ *lacI*⁺ pACG_araBAD_ *geneX* mutant strain under different L-arabinose concentrations and cultivation modes are shown in Figure 37 and Figure 38, respectively.

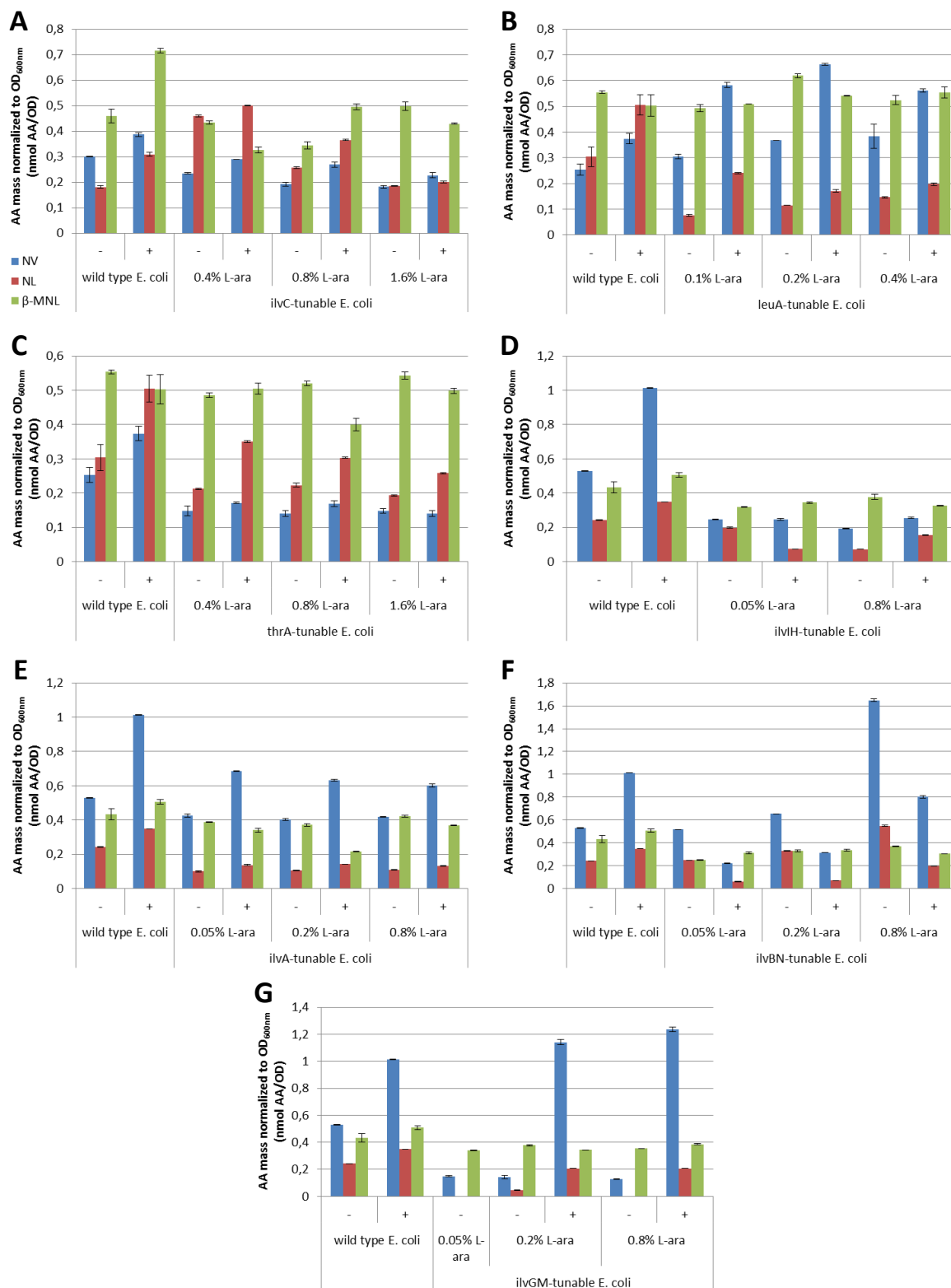


Figure 37. Molar concentrations of norvaline (blue bars), norleucine (red bars) and β-methylnorleucine (green bars) normalized to OD_{600nm} in the intracellular soluble protein fraction from samples taken 3h after IPTG induction of cultivations of *ilvC*-tunable *E. coli* (A), *leuA*-tunable *E. coli* (B), *thrA*-tunable *E. coli* (C), *ilvIH*-tunable *E. coli* (D), *ilvA*-tunable *E. coli* (E), *ilvBN*-tunable *E. coli* (F) and *ilvGM*-tunable *E. coli* (G) in a 10mL mini-reactor subjected to different L-arabinose concentrations and cultivation modes. Two cultivation modes were tested: standard cultivation conditions (-) and conditions triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+). Strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ (indicated as “wild type *E. coli*” in the chart) was employed as a control for comparison.

Under standard cultivation conditions, for *ilvC*- and *ilvIH*-tunable *E. coli* strains, norvaline and norleucine concentrations in the intracellular soluble protein fraction decreased when adding increasing concentrations of L-arabinose into the medium. The opposite behavior was observed for *ilvBN*-tunable *E. coli*. For *thrA*-, *ilvA*- and *ilvGM*-tunable *E. coli* strains, norvaline and norleucine concentrations showed a reduction with respect to the control *E. coli* strain but L-arabinose concentration did not seem to have a clear effect on ncBCAA concentration (Figure 37,-). The highest reduction of norvaline concentration with respect to the control *E. coli* strain was reported when inducing *ilvIH*- (-63.6%) and *ilvGM*-tunable *E. coli* strains (-75.9%) with 0.8 % L-arabinose. The highest decrease of norleucine concentration was shown after inducing *ilvIH*- (-69.9%) and *ilvGM*-tunable *E. coli* strains (-100%) with 0.8 % L-arabinose. The value of 100% did not correspond to a measured norleucine concentration of 0. In that case norleucine concentration could simply not be measured since it was lower than the detection limit of the GC.

Under standard cultivation conditions, for almost all tested strains and L-arabinose concentrations, no significant variation was reported for β -methylnorleucine concentration in the intracellular soluble protein fraction, with exception of *ilvBN*-tunable *E. coli* induced with 0.05 % L-ara, which shown a reduction around 42%. In addition, effect of increasing L-arabinose concentrations did not have a clear effect on β -methylnorleucine concentration (Figure 37,-).

Under cultivation conditions triggering ncBCAA formation (i.e. pyruvate pulses combined with O₂ limitation), for *ilvC*- and *leuA*-tunable *E. coli* strains, norvaline and norleucine concentration in the intracellular soluble protein fraction tended to decrease when adding increasing concentrations of L-arabinose into the medium. The opposite behavior was observed for *ilvBN*- and *ilvIH*-tunable *E. coli* strains. For *thrA*- and *ilvA*-tunable *E. coli* strains, norvaline and norleucine concentration showed a reduction with respect to the control *E. coli* strain but L-arabinose concentration did not seem to have a clear effect on ncBCAA concentration. For *ilvGM*-tunable *E. coli*, only norleucine concentration reported a reduction with respect to the control *E. coli* strain. L-arabinose concentration did not seem to have a clear effect on ncBCAA concentration in this case (Figure 37,+).

For the intracellular soluble protein fraction, similar results were obtained when expressing data as ratio of the concentration of ncBCAA with respect to the canonical counterpart (Figure S74).

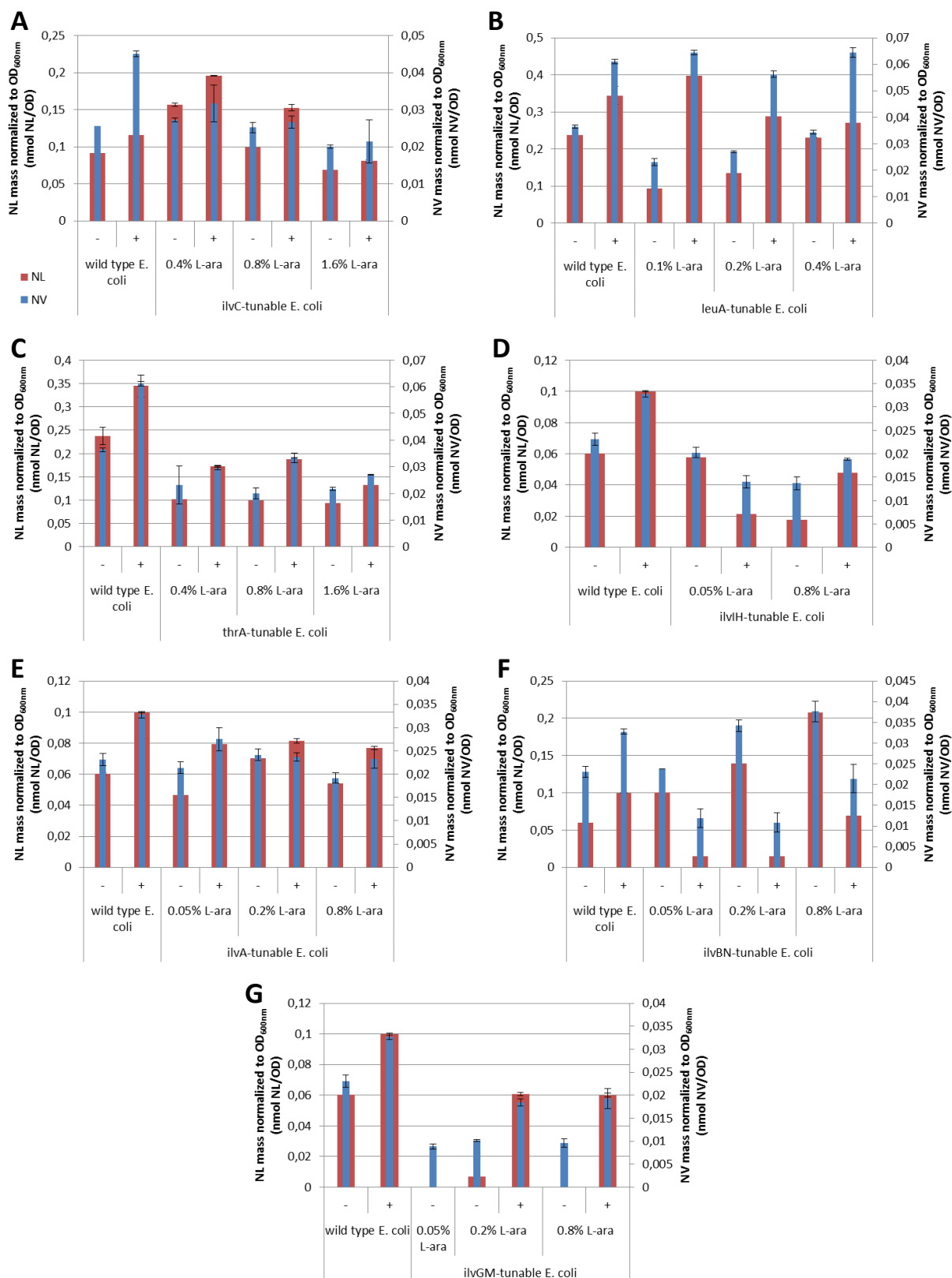


Figure 38. Molar concentrations of norvaline (blue bars) and norleucine (red bars) normalized to OD_{600nm} in the inclusion body fraction from samples taken 3 h after IPTG induction of cultivations of *ilvC*-tunable *E. coli* (A), *leuA*-tunable *E. coli* (B), *thrA*-tunable *E. coli* (C), *ilvIH*-tunable *E. coli* (D), *ilvA*-tunable *E. coli* (E), *ilvBN*-tunable *E. coli* (F) and *ilvGM*-tunable *E. coli* (G) in a 10mL mini-reactor subjected to different L-arabinose concentrations and cultivation modes. Two cultivation modes were tested: standard cultivation conditions (-) and conditions

triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+). Strain *E. coli* K-12 BW25113 pSW3_ *lact*⁺ (indicated as “wild type *E. coli*” in the chart) was employed as a control for comparison.

Under standard cultivation conditions, *ilvC*- and *ilvIH*-tunable *E. coli* reported a significant decrease of norvaline and norleucine concentrations in the inclusion body fraction when adding increasing concentrations of L-arabinose into the medium. The opposite behavior was observed for *ilvBN*-tunable *E. coli*. For *thrA*- and *ilvGM*-tunable *E. coli* strains, norvaline and norleucine concentrations showed a significant reduction with respect to the control *E. coli* strain but effect of increasing L-arabinose concentrations did not show a clear effect on ncBCAA concentration. For *ilvA*-tunable *E. coli* no significant variation of norvaline and norleucine concentration was reported with respect to the control *E. coli* strain and effect of increasing L-arabinose concentrations did not seem to show a clear effect on ncBCAA concentration (Figure 38,-). The highest reduction of norvaline concentration was reported when inducing *ilvIH*-tunable *E. coli* with 0.8 % L-arabinose (-40.7%) and *ilvGM*-tunable *E. coli* with 0.05% L-arabinose (-61.8%). The highest decrease of norleucine concentration was shown after inducing *ilvIH*- (-70.8%) and *ilvGM*-tunable *E. coli* (-100%) with 0.8 % L-arabinose. As mentioned before, the value of 100% did not correspond to a measured norleucine concentration of 0. In that case norleucine concentration could simply not be measured since it was lower than the detection limit of the GC.

Under cultivation conditions triggering ncBCAA formation (i.e. pyruvate pulsing combined with oxygen limitation), all tested strains with exception of *leuA*- and *ilvIH*-tunable *E. coli* showed the same pattern described at the previous paragraph. Interestingly, the two aforementioned strains reported the opposite effect than under standard cultivation conditions (Figure 38,+).

For the inclusion body fraction, similar results were obtained when expressing data as ratio of the concentration of ncBCAA with respect to the canonical counterpart (Figure S75).

5.9 Screening of potential *ilvGM*- and *ilvIH*-tunable *E. coli* strains in a 15 L reactor under conditions triggering ncBCAA formation

According to section 5.8 strains *E. coli* K-12 BW25113 pSW3_*lacI*⁺ pACG_araBAD_*ilvGM* (*ilvGM*-tunable *E. coli*) and *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_*lacI*⁺ pACG_araBAD_*ilvIH* (*ilvIH*-tunable *E. coli*) induced with 0.8% L-arabinose showed the best performance among all screened mutants in a 10 mL mini-reactor, since they reported the most significant reduction of ncBCAA mis-incorporation into recombinant mini-proinsulin in comparison with the control non-engineered *E. coli* strain. The aim of this experiment was to verify the performance of the aforementioned potential tunable *E. coli* strains in a 15L reactor under cultivation conditions triggering formation of ncBCAA, i.e. pyruvate pulses and oxygen limitation, in order to confirm its advantage as strain ensuring product quality. For comparison, the control non-engineered *E. coli* host (*E. coli* K-12 BW25113 pSW3_*lacI*⁺) was also cultivated.

5.9.1 Cultivation mode

5.9.1.1 Cultivation of *E. coli* K-12 BW25113 pSW3_*lacI*⁺ (control strain) under conditions triggering ncBCAA formation

Cultivation operation was already described in section 5.7.2.1.2 (Figure 30).

5.9.1.2 Cultivation of *ilvGM*-tunable *E. coli* under conditions triggering ncBCAA formation

Cultivation operation was already described in section 5.7.2.1.2 (Figure 30) and only minor changes were done in order to adapt the cultivation process to *ilvGM*-tunable *E. coli* strain. Both pre-culture and reactor medium contained additionally 25 µg/mL chloramphenicol. The reactor medium additionally contained 0.8% L-arabinose, necessary to induce expression of gene *ilvGM* hosted in plasmid pACG_araBAD_*ilvGM*. The feeding solution was also additionally supplemented with 25 µg/mL chloramphenicol and 0.8% L-arabinose. A general overview of the cultivation is shown in Figure 39.

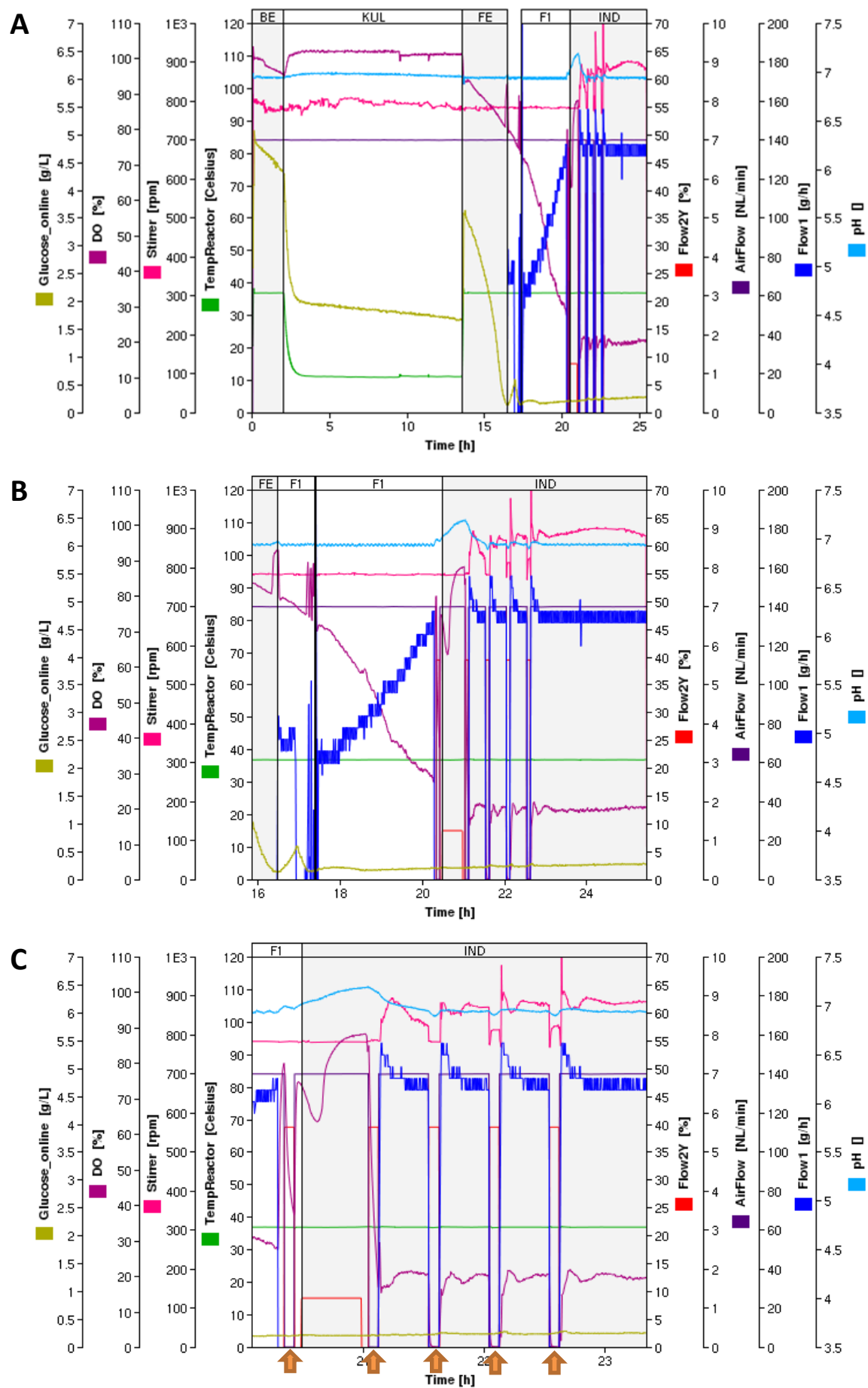


Figure 39. Overview of the cultivation of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* (*ilvGM*-tunable *E. coli*) in a 15L reactor under conditions triggering ncBCAA formation, during the whole cultivation process (A), during fed-batch period (B) and during pyruvate pulsing (C). Different cultivation phases are shown in the diagram as *BE* (first 2h of batch phase), *KUL* (12h cold period at 15 °C), *FE* (remaining 2h of batch phase), *F1* (3h exponential fed-batch phase) and *IND* (induction, linear fed-batch phase and pyruvate pulsing). IPTG induction was performed for 30 minutes (20.5 to 21 h cultivation time). Pyruvate pulses are indicated by orange arrows (C). Present in the diagram axes, *Flow1* corresponds to the flow rate (g/h) of the pump transporting the feed solution into the reactor while *Flow2Y* corresponds to one tenth of the flow rate (g/h) of the pump transporting either the IPTG solution used for induction or the pyruvate solution for pulsing. Exponential fed-batch was, due to a programming error, started shortly before glucose was completely consumed in the batch phase. Hence a small glucose accumulation was reported at around 17h. Thus, exponential feeding was shortly shut down until glucose was completely depleted and then, activated again.

5.9.1.3 Cultivation of *ilvIH*-tunable *E. coli* under conditions triggering ncBCAA formation

Cultivation operation was already described in section 5.7.2.1.2 (Figure 30) and only minor changes were done in order to adapt the cultivation process to *ilvIH*-tunable *E. coli*. Both pre-culture and reactor medium contained additionally 25 µg/mL chloramphenicol. The reactor medium additionally contained 0.8% L-arabinose, necessary to induce expression of gene *ilvIH* hosted in plasmid pACG_araBAD_ *ilvIH*. The feeding solution was also additionally supplemented with 25 µg/mL chloramphenicol and 0.8% L-arabinose. A general overview of the cultivation is shown in Figure 40.

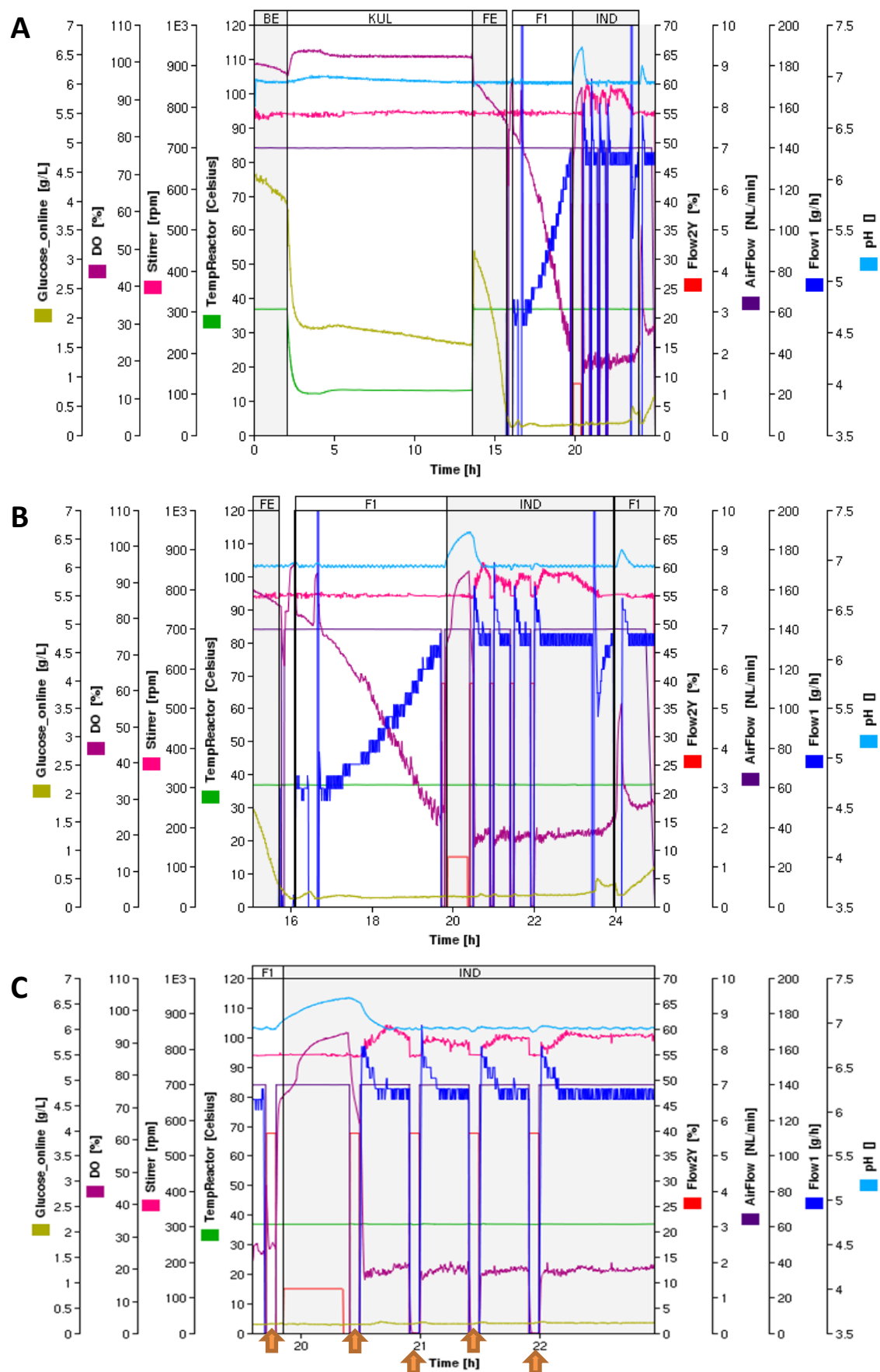


Figure 40. Overview of the cultivation of *E. coli* K-12 BW25113 $\Delta ilvIH$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH* (*ilvIH*-tunable *E. coli*) in a 15L reactor under conditions triggering ncBCAA formation, during the whole cultivation process (A), during fed-batch period (B) and during pyruvate pulsing (C). Different cultivation phases are shown in the diagram as *BE* (first 2h of batch phase), *KUL* (12h cold period at 15 °C), *FE* (remaining 2h of batch phase), *F1* (3h exponential fed-batch phase) and *IND* (induction, linear fed-batch phase and pyruvate pulsing). IPTG induction was performed for 30 minutes (19.9 to 20.4 h cultivation time). Pyruvate pulses are indicated by orange arrows (C). Present in the diagram axes, *Flow1* corresponds to the flow rate (g/h) of the pump transporting the feed solution into the reactor while *Flow2Y* corresponds to one tenth of the flow rate (g/h) of the pump transporting either the IPTG solution used for induction or the pyruvate solution for pulsing. Exponential fed-batch was, due to a programming error, started shortly before glucose was completely consumed in the batch phase. Hence a small glucose accumulation was reported at around 16.5hh. Thus, exponential feeding was shortly shut down until glucose was completely depleted and then, activated again. At around 23.5h glucose suddenly started being accumulated. It was hypothesized that the pump was going too fast even though *Flow1* values were correct. Hence, pump was shortly shut down and restarted. However, after reaching the pump again the set *Flow1* value, glucose still kept accumulating. This might be explained due to a problem inherent to the cultivated strain, i.e. culture is reaching dead phase.

5.9.2 Mini-proinsulin analysis by HPLC

Recombinant mini-proinsulin concentration from hourly samples was analyzed according to an *in house* HPLC method available at Sanofi-Aventis Deutschland GmbH. 1 mL of culture broth was directly used for analysis. PPI concentrations resulting from the *in house* HPLC analytical method are shown in Figure 41. PPI concentrations determined for all three tested *E. coli* strains were really similar. In addition, cell growth behavior (represented by OD_{600nm} and CDW) was also comparable for all tested strains (Figure S76), suggesting that specific production of recombinant mini-proinsulin remains the same, independently of the strain employed.

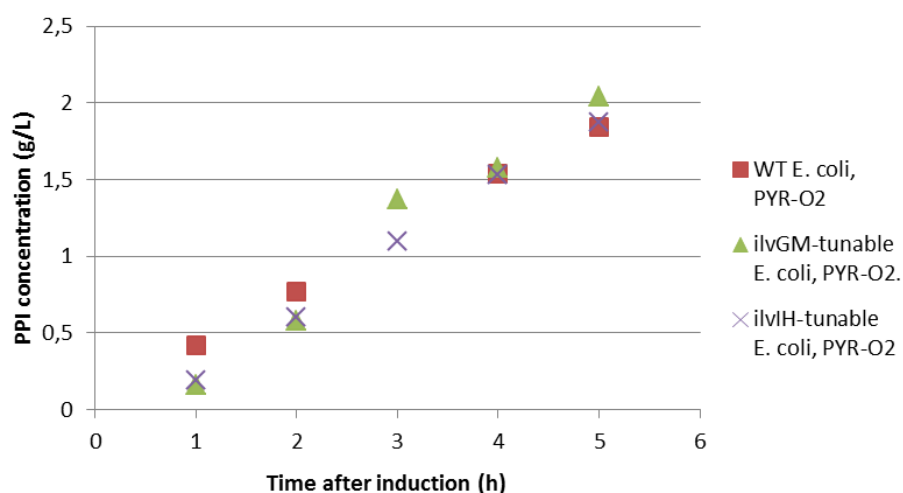


Figure 41. Estimated recombinant mini-proinsulin concentrations over time after induction of different *E. coli* cultivations in a 15 L reactor under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2). Indicated in the legend, “WT *E. coli*” refers to the wild type strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺, “*ilvGM*-tunable *E. coli*” alludes to strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* and “*ilvIH*-tunable *E. coli*” corresponds with strain *E. coli* K-12 BW25113 $\Delta ilvIH$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH*.

5.9.3 Analysis of ncBCAA

Samples for amino acid analysis by GC-FID were prepared exactly as aforementioned in section 5.7.2.3. Concentrations of ncBCAA present in the intracellular soluble protein fraction and in the inclusion body fraction over cultivation time for each tested strain are shown in Figure 42 and Figure 43, respectively.

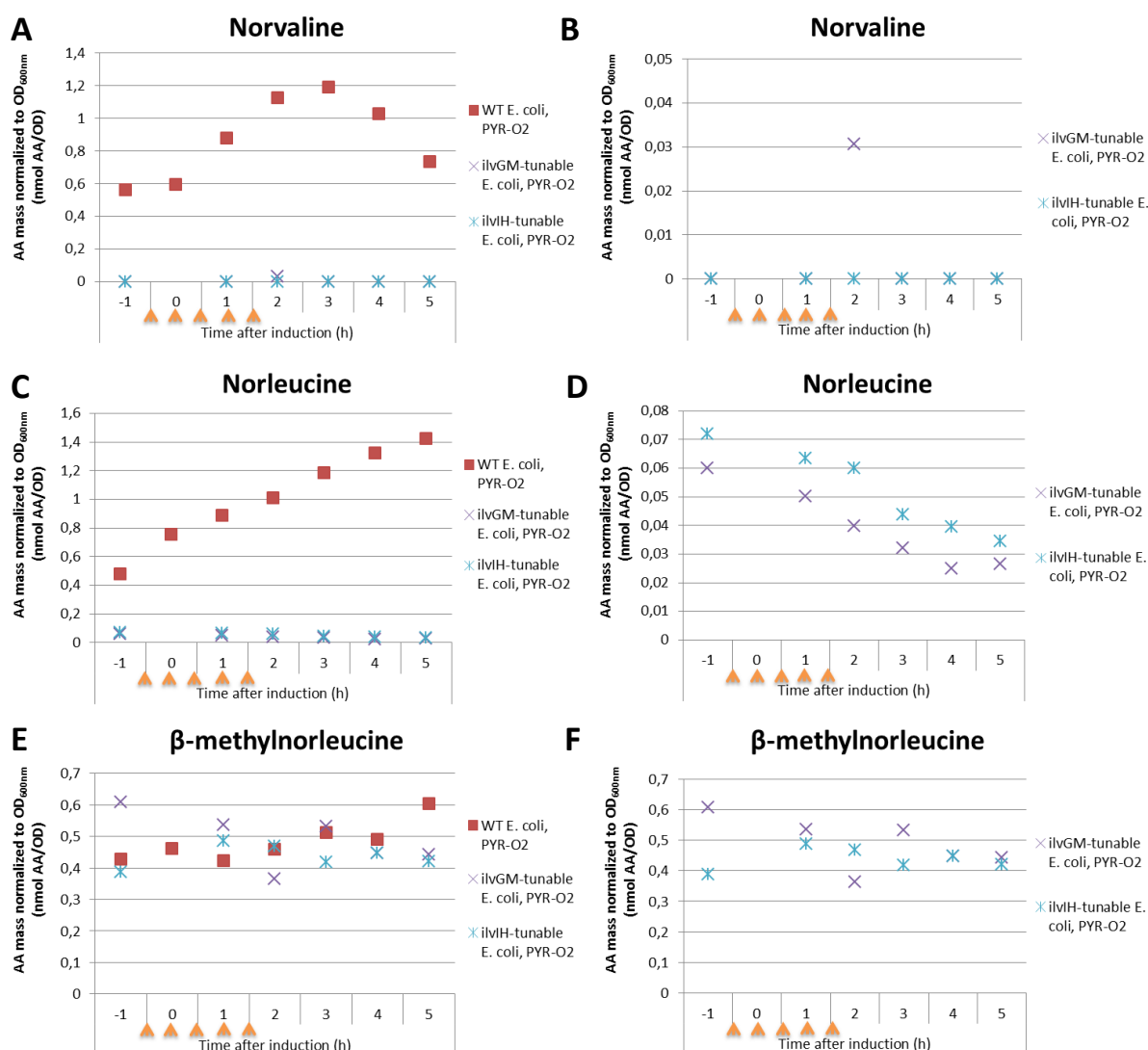


Figure 42. Molar concentrations of norvaline (A, B), norleucine (C, D) and β -methylnorleucine (E, F) normalized to OD_{600nm} present in the intracellular soluble protein fraction calculated over time after induction of different *E. coli* cultivations in a 15L reactor under cultivation conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2), in a big (A, C, E) and reduced y-axis scale (B, D, F). Indicated in the legend, “WT *E. coli*” refers to the wild type strain *E. coli* K-12 BW25113 pSW3_ *lact*⁺; “*ilvGM-tunable E. coli*” alludes to strain *E. coli* K-12 BW25113 pSW3_ *lact*⁺ pACG_araBAD_ *ilvGM* and “*ilvIH-tunable E. coli*” corresponds with strain *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_ *lact*⁺ pACG_araBAD_ *ilvIH*. Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O₂ limitation was applied.

The cultivation of the control *E. coli* strain subjected to pyruvate pulses combined with O₂ limitation (“WT *E. coli*, PYR-O2”) reported a progressive accumulation of norleucine and β -methylnorleucine in

the intracellular soluble protein fraction over time after induction, being that more significant for norleucine. Furthermore, norvaline concentration also increased progressively under aforementioned cultivation conditions, but only until 3h after induction. From that time point on, norvaline concentration progressively dropped until reaching initial values at 5 h after induction. This might suggest that, after 2h from last pyruvate pulse combined with O₂ limitation, its associated effect triggering norvaline accumulation is not active anymore (Figure 42).

Both tested potential mutants in cultivations “*ilvGM*-tunable *E. coli*, PYR-O2” and “*ilvIH*-tunable *E. coli*, PYR-O2” reported a dramatic reduction of norvaline and norleucine concentrations in the intracellular soluble protein fraction, being such decrease higher for norleucine in “*ilvGM*-tunable *E. coli*, PYR-O2”. However, β -methylnorleucine concentrations did not significantly vary with respect to the control cultivation. It is noteworthy to highlight that, for most samples, norvaline could not be properly detected since concentrations were under the limit of detection of the GC-FID equipment (Figure 42).

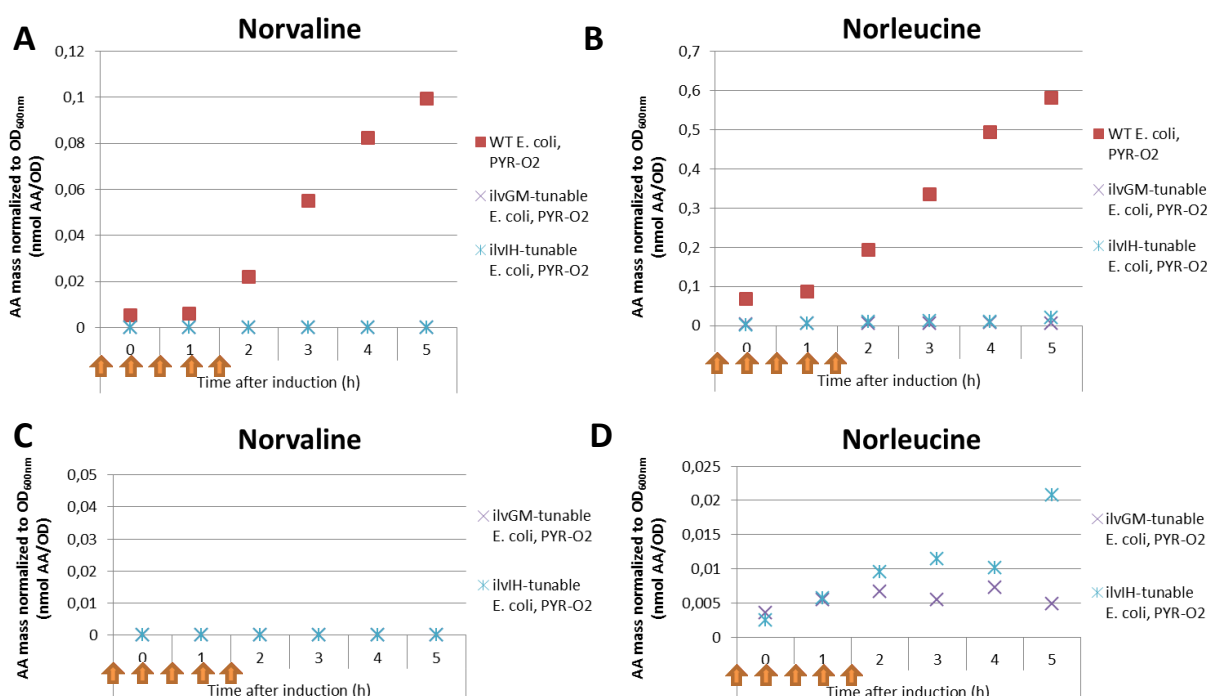


Figure 43. Molar concentrations of norvaline (A, C) and norleucine (B, D) normalized to OD_{600nm} present in the inclusion body fraction calculated over time after induction of different *E. coli* cultivations in a 15L reactor under cultivation conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O2), in a big (A, B) and reduced y-axis scale (C, D). Indicated in the legend, “WT *E. coli*” refers to the wild type strain *E. coli* K-12 BW25113 pSW3_*lacI*⁺. “*ilvGM*-tunable *E. coli*” alludes to strain *E. coli* K-12 BW25113 pSW3_*lacI*⁺ pACG_araBAD_*ilvGM* and “*ilvIH*-tunable *E. coli*” corresponds with strain *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_*lacI*⁺ pACG_araBAD_*ilvIH*. Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O₂ limitation was applied.

The cultivation of the control *E. coli* strain subjected to pyruvate pulses combined with O₂ limitation (“WT *E. coli*, PYR-O2”) reported a progressive accumulation of norvaline and norleucine in the inclusion body fraction over time after induction. Again, and similar to reported in the intracellular soluble fraction, both tested potential mutants in cultivations “*ilvGM*-tunable *E. coli*, PYR-O2” and “*ilvIH*-tunable *E. coli*, PYR-O2” reported a dramatic reduction of norvaline and norleucine

concentrations in the inclusion body fraction, being this decrease even higher for norleucine in “*ilvGM*-tunable *E. coli*, PYR-O2”. Norvaline could not be detected in any case for both tested mutants. β -methylnorleucine could not be detected in any tested samples (Figure 43).

Similar results are also shown for the inclusion body fraction when normalizing amino acid concentration to PPI mass (Figure S77).

5.9.4 Acetate and formate analysis

Acetate and formate concentrations from hourly samples were offline analysed by *in house* enzymatic assays available at Sanofi-Aventis Deutschland GmbH. Sample preparation for acetate and formate analysis was as follows: 10 mL of culture broth were inactivated with Bardac, centrifuged at 4000 g for 10 min and resultant supernatant was decanted and filtered through a 0.45 μ m filter. The filtrate was then used for analysis. Estimated concentrations of acetate and formate in culture broth supernatants over cultivation time are shown in Figure 44.

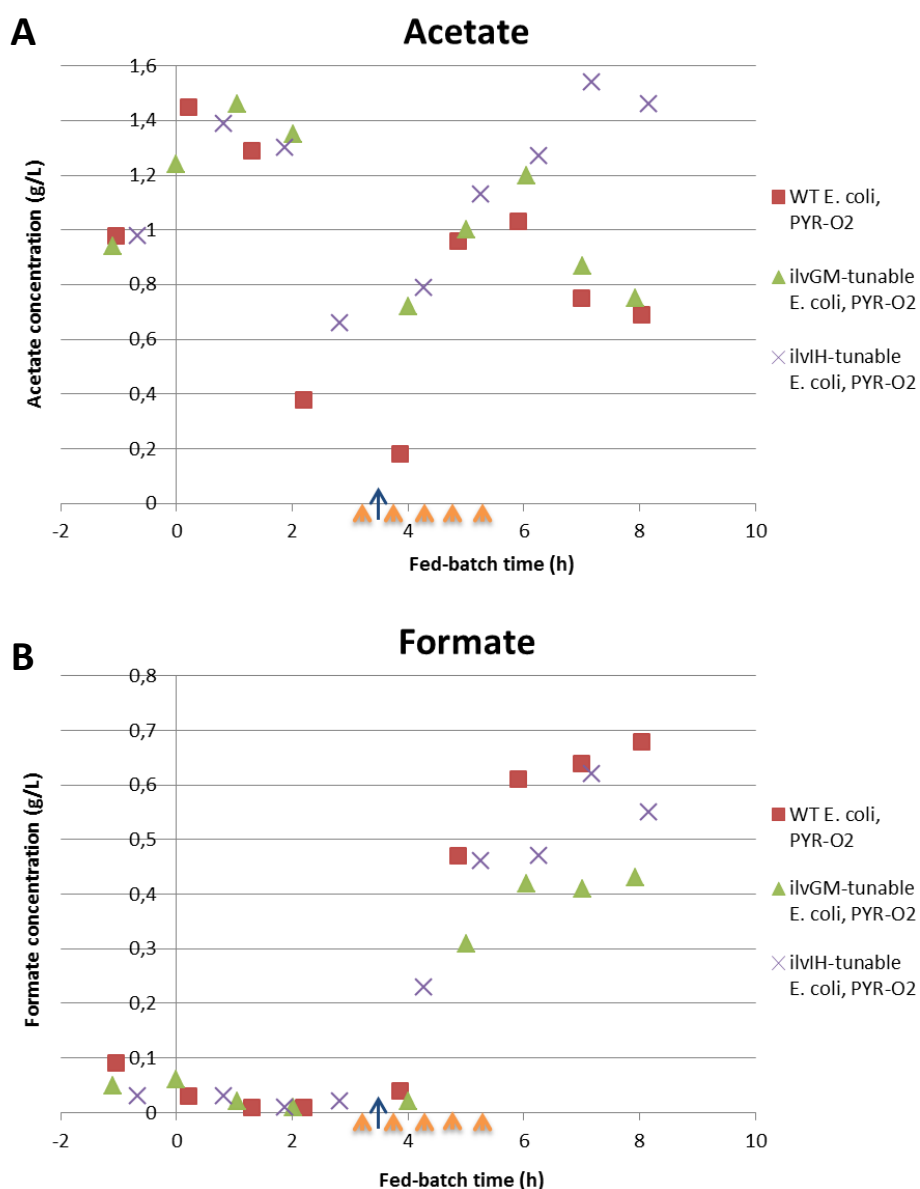


Figure 44. Concentration of acetate (A) and formate (B) present in culture broth supernatants calculated over fed-batch time of different *E. coli* cultivations in a 15L reactor under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O₂). Indicated in the legend, “WT *E. coli*” refers to the wild type strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺, “*ilvGM*-tunable *E. coli*” alludes to strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* and “*ilvIH*-tunable *E. coli*” corresponds with strain *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH*. Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O₂ limitation was applied. Blue arrow points out time point where IPTG induction was carried out.

For all tested cultivations, a similar acetate profile was observed over time. Acetate concentration reached a maximum after batch phase was finished (0 h fed-batch time). Afterwards, acetate levels decreased to a minimum around 0.2 g/L at about 4h after induction. From that time on, acetate concentration rapidly started to increase as a consequence of the metabolic alteration triggered by pyruvate pulsing combined with O₂ limitation. Once pulsing was finished, acetate levels started to slowly decrease until the end of the cultivation, except for cultivation “*ilvIH*-tunable *E. coli*, PYR-O₂”, where acetate still increased (Figure 44). This is due to the problems of glucose accumulation reported at the end of this cultivation (Figure 40).

For all tested cultivations, a similar formate profile was observed over time as well. Formate levels remained close to 0 until about 4h after induction. However, from that time point on, formate concentration progressively increased as a consequence of the metabolic alteration triggered by pyruvate pulsing combined with O₂ limitation. Once pulsing was finished, formate levels stopped increasing, reaching a plateau (Figure 44).

6. Discussion

6.1 Analysis of mini-proinsulin expression in *E. coli* K-12 BW25113 containing different variants of plasmid pSW3

Expression of mini-proinsulin from plasmid pSW3 is under the control of a P_{tac} promoter. P_{tac} is a hybrid promoter where -35 region is derived from the trp promoter while -10 region comes from the lac UV5 promoter (de Boer *et al.*, 1983). The P_{tac} promoter is inducible by IPTG addition but, under certain conditions, it is also leaky, i.e. expression is also triggered in a basal level even when no inducer is present. This is because the LacI repressor does not tightly bind to the operator and can dissociate. Hence, when using *lac*-based promoters in cloning vectors is important to take into consideration that enough LacI repressor is present in order to avoid leaky expression of the promoter. The number of LacI molecules necessary to repress a *lac*-based promoter mainly depends on the number of operators regulating the promoter, the affinity of the repressor to the operator and the copy number of the plasmid bearing the *lac* promoter (Penumetcha *et al.*, 2010). Different variants of *lacI* promoter sequences leading to different expression levels of *lacI* have been reported in *E. coli* strains: *lacI*⁺ and *lacI*^q in Calos (1978), I-UJ177 in Calos and Miller (1980) and *lacI*^{q1} in Calos and Miller (1981). According to Glascock *et al.* (1998), an *E. coli* wild type cell (*lacI*⁺ variant) has about 10 LacI molecules. The *lacI*^q promoter version shows 10-fold expression enhancement of *lacI* if compared with the *lacI*⁺ variant (100 LacI molecules) while expression by *lacI*^{q1} is 17-fold stronger than that of *lacI*^q and 170-fold stronger than that of *lacI*⁺ (1700 LacI molecules).

Plasmid pSW3 does not contain *lacI* and it was especially designed for expression in strain *E. coli* K-12 W3110M. This strain is *lacI*^q and it expresses endogenously 10-fold more LacI repressor than a *lacI*⁺ strain, hence allowing a tight expression of plasmid pSW3. However, in this study, *E. coli* K-12 BW25113 was used as a model strain. Results reported in section 5.1 revealed that, as opposed to *E. coli* K-12 W3110M, P_{tac} promoter controlling expression of mini-proinsulin in plasmid pSW3 is leaky in the genetic background of *E. coli* K-12 BW25113, being no significant difference in mini-proinsulin expression level between induced and non-induced samples. This suggests that *E. coli* K-12 BW25113 is a *lacI*⁺ strain so that there is not sufficient endogenous LacI repressor to maintain P_{tac} promoter inactive when no IPTG induction is performed. However, plenty of literature wrongly stated that *E. coli* K-12 BW25113 was *lacI*^q (reviewed by Baba *et al.*, 2006, Supplementary table 1). Moreover, the sequence of the 3 allelic variants of the *lacI* promoter (*lacI*⁺, *lacI*^q and *lacI*^{q1}) were aligned to the *lacI* promoter located in the genome of *E. coli* K-12 BW25113 and it was confirmed that strain is actually *lacI*⁺ (Figure S78).

Observed results in section 5.1 are in accordance with the investigation of Glascock and Weickert (1998), where a similar scenario was described. They analyzed the recombinant protein expression behavior of the *lacI*⁺ *E. coli* strain SGE1661 containing the medium-copy plasmid pSGE714 (no *lacI*), which expresses recombinant β -galactosidase under the control of a P_{tac} promoter. They showed that non-induced cells already express around half of the total recombinant β -galactosidase produced by cells induced with 0.1 mM IPTG. However, it is noteworthy to point out that: (i) unlike plasmid pSGE714, plasmid pSW3 is a high-copy plasmid, since it contains the pBR22 origin of replication

without the *rop* gene; (ii) instead of the 0.1 mM IPTG used in Glascock and Weickert (1998), 0.5 mM IPTG was used in the current study. These two considerations would explain why no significant differences in mini-proinsulin expression levels were reported between induced and non-induced samples.

An *E. coli* K-12 BW25113 cell, as *lacI*⁺, would have approximately 10 LacI molecules, which is not sufficient to repress P_{tac} promoter present in pSW3 because there are 115 plasmid copies per cell (or 115 pSW3 operator sites per cell, as there is a single operator in the P_{tac} promoter) (Glascock and Weickert, 1998). An *E. coli* K-12 W3110M cell, as *lacI*^q, would have approximately 100 LacI molecules and it should be sufficient to repress the P_{tac} promoter. Considering that strain *E. coli* K-12 BW25113 (*lacI*⁺) expressing a pSW3 variant contains about 115 copies of the plasmid, expression levels of *lacI* are predicted to be 0, 1150 and 11500 LacI molecules for pSW3, pSW3_*lacI*⁺ and pSW3_*lacI*^q, respectively (Table 28). Those estimations are in accordance with results reported in section 5.2. As opposed to plasmid pSW3, plasmid variants pSW3_*lacI*⁺ and pSW3_*lacI*^q did not report promoter leakiness, confirming sufficient LacI levels. However, weak expression was shown for plasmid pSW3_*lacI*^q, which could be explained due to the high content of LacI repressor expressed. Hence, it may be that the concentration of IPTG used for induction in that specific case was not high enough to effectively trigger expression of the recombinant protein. It might also be that the high amount of LacI being expressed reduced the expression efficiency of the recombinant protein due to an overexploitation of the transcription and translation cell machinery.

All in all, research hypothesis 7 could be confirmed.

Table 28. Estimated number of LacI molecules expressed by two different *E. coli* K-12 strains expressing different pSW3 plasmid variants. Data obtained from Glascock and Weickert (1998).

Strain	Estimated number of LacI molecules expressed in the plasmid (115 copies/cell)	Estimated number of LacI molecules expressed in the genome
<i>E. coli</i> K-12 W3110M (<i>lacI</i> ^q)	0	100
<i>E. coli</i> K-12 W3110M (<i>lacI</i> ^q) pSW3	0	100
<i>E. coli</i> K-12 BW25113 (<i>lacI</i> ⁺)	0	10
<i>E. coli</i> K-12 BW25113 (<i>lacI</i> ⁺) pSW3	0	10
<i>E. coli</i> K-12 BW25113 (<i>lacI</i> ⁺) pSW3_ <i>lacI</i> ⁺	1150	10
<i>E. coli</i> K-12 BW25113 (<i>lacI</i> ⁺) pSW3_ <i>lacI</i> ^q	11500	10

6.2 Evaluation of L-arabinose induction in *E. coli* BW25113 Δ *geneX* expressing pSW3_*lacI*⁺ and pACG_araBAD_*geneX* (*geneX*-tunable *E. coli*)

After screening of various L-arabinose concentrations, the effective induction range of L-arabinose for plasmid pACG_araBAD_*geneX* was determined to be 0.05-1.6 % (section 5.3). This concentration range is higher than the one normally used for commercial plasmids based in the *araBAD* promoter.

For instance, induction of the commercial plasmid pBAD_DEST49 is recommended within the range 0.00002-0.2 % L-arabinose (Website 4). Variation in the L-arabinose amounts necessary for induction might be explained due to differences in the selected origin of replication, which determines the plasmid copy number in the cell. While plasmid pBAD_DEST49 contains a pUC origin, acting as a high copy plasmid, plasmid pACG_araBAD_geneX contains an F plasmid-based origin of replication, ensuring 1-copy plasmid. The selection of araBAD promoter variants with enhanced transcriptional strength (Website 5) would have probably resulted in a reduction of the effective L-arabinose induction range.

According to results reported in section 5.3, it is noteworthy that, for each target gene, different induction strength, i.e. L-arabinose concentration, was necessary in order to trigger genetic expression levels enough to recover cell growth levels of the control *E. coli* strain. This might be explained due to the different degree of essentiality of the target genes for the cell. The L-arabinose concentration needed to fully recover growth levels of the control *E. coli* strain was especially high for the *thrA*-tunable *E. coli* strain (0.4 %), followed by *ilvC*-tunable *E. coli* (0.4 %), *leuA*-tunable *E. coli* (0.1 %) and *ilvA*-tunable *E. coli* (0.05 %). Moreover, when no L-arabinose was supplemented, cell growth reported for the aforementioned *E. coli* mutants was close to 0. These results are in accordance with the study of Baba *et al.* (2006). They generated numerous single gene knock-out *E. coli* mutants and then tested growth behavior (OD_{600nm}) of the different mutants 24 and 48 h after cultivation in LB and minimal medium (MOPS). In that study, the aforementioned genes were classified as non-essential since mutants were able to grow in LB medium. However, growth in minimal medium was reported to be almost 0.

For *ilvIH*-, *ilvBN*- and *ilvGM*-tunable *E. coli* strains, no significant differences in growth behavior between non-induced and induced samples were reported, independently on the L-arabinose concentration employed for induction (section 5.3). As a consequence, differently from reported for mutants mentioned in the previous paragraph, induction efficiency of L-arabinose could not be evaluated by monitoring cell growth. These results suggest that lack of one of the three *ilvGM*, *ilvIH* and *ilvBN* gene products does not dramatically affect growth behavior since, as isoenzymes, its function can be also carried out by the other two available gene products. These results are also in accordance with the study of Baba *et al.*, (2006). They showed that *ilvIH*-, *ilvBN*- and *ilvGM* mutants were also able to grow in minimal medium.

6.3 Evaluation of m-toluate induction in *E. coli* BW25113 Δ*geneX* expressing pSW3_ *lacI*⁺ and pACG_XylSPm_ *geneX* (*geneX*-tunable *E. coli*)

As demonstrated in section 5.4, addition of m-toluic acid does not have an effect on growth behavior if compared with the non-induced case for any tested strains. This might be explained due to non-sufficient induction strength of the selected XylSPm/Pm promoter variant. The unmodified wild type Pm promoter variant was selected for this study since it was demonstrated to have a less basal expression than other mutagenized high-level expression variants (Balzer *et al.*, 2013; Binder *et al.*, 2016). However, this promoter variant was successfully tested mainly in medium copy plasmids (40-60 copies per cell) containing the RK2 origin of replication (i. e. plasmid pJB658, Blatny *et al.*, 1997). In the current study, however, the selected plasmid for tunable expression contains an F plasmid-based origin of replication, ensuring 1-copy plasmid, and this might explain the non-sufficient

induction of the Pm promoter. The generation of new pACG_XylSPm_geneX plasmid variants containing other mutagenized high-level expression variants of the Pm promoter (Winther-Larsen *et al.*, 2000) might improve the induction strength. Nevertheless, since induction of the tunable plasmid pACG_araBAD_geneX was demonstrated to properly work, this variant was selected for gene regulation and no further investigation was performed with the pACG_XylSPm_geneX plasmid variant. In addition, it can be concluded that 10 mM m-toluic acid triggers cellular toxicity since under this concentration OD_{600nm} reached almost 0 for all tested strains. This is not in accordance with many literature reports stating as an advantage of the XylSPm/Pm promoter system the non-toxic character of m-toluic acid for *E. coli* cells (Binder *et al.*, 2016).

6.4 Establishment of a GC-FID method allowing analysis of canonical and non-canonical amino acids

Among the analyzed amino acids, β -methylnorleucine was the only one represented by two different peaks in the chromatogram (Table 24, section 5.5). Since β -methylnorleucine contains two asymmetric carbons, 4 different configurations, i.e. 4 stereoisomers, are possible (SS, SR, RS and RR). Configurations SS and RR and configurations SR and RS are enantiomers and have the same chemical and physical properties. Hence, when injecting a β -methylnorleucine solution containing the 4 stereoisomers into the GC for analysis, two peaks are generated due to variations in the retention time of each pair of enantiomers. The β -methylnorleucine standard employed for calibration contained the four variants. However, according to Muramatsu *et al.* (2002), the final stereochemical configuration of the β -methylnorleucine produced by *E. coli* is predicted to be 2S,3S because this is also the configuration of the natural L-isoleucine biosynthesized by *E. coli*. Hence, only the first peak at 3.398 min was considered for β -methylnorleucine calibration and analysis.

Concerning the evaluation of acid hydrolysis effect on amino acid analysis (Figure 18, section 5.5), a signal reduction was reported after hydrolysis for most of the analyzed amino acids, being this especially evident for norvaline, norleucine, threonine, methionine and tyrosine. In addition, amino acids asparagine, glutamine and tryptophan could not be measured after hydrolysis and the signal of glutamate and aspartate reported a significant increase. These results are in accordance with previous investigations such as the one from Pickering and Newton (1990) and the one from Davidson (1997). According to Davidson (1997), serine and threonine are recovered in low yield due to ester formation and due to the modification by dehydration of the hydroxyl group present in the side chain. Tyrosine is also generated in lower yields after hydrolysis due to modification of the phenolic group in the side chain. Methionine is as well obtained in low yields after hydrolysis due to oxidation of the thioether group present in the side chain. Tryptophan is not quantifiable since the indole group of the molecule is destroyed. Asparagin and glutamin are deamidated to the respective acids (aspartate and glutamate). In addition, recovery of some hydrophobic amino acids such as valine, isoleucine, leucine and alanine is also poor due to the high stability of the bonds generated between those mentioned amino acids.

Depending on the target amino acid, different strategies have been described in order to diminish effects of acid hydrolysis in the amino acid recovery yield: addition of reducing agents (for methionine and tryptophane), drying of hydrolysates (for serine and threonine), addition of phenol (for tyrosine) and/or extension of hydrolysis time (for valine, isoleucine, leucine and alanine), among

others (Davidson, 1997). In this study, it was demonstrated that the extension of acid hydrolysis from 24h to 72h slightly improved recovery yield of most of the analyzed amino acids, reaching values close to the theoretical ones (Figure 19, section 5.5). However, due to practical reasons and the fact that improvement of hydrolysis performance was not dramatic, 24h was finally selected as the standard time for acid hydrolysis in this study.

6.5 Establishment of cultivation conditions based on pyruvate pulsing and O₂ limitation leading to an increase of ncBCAA mis-incorporation into recombinant mini-proinsulin in *E. coli*

During fermentation in large-scale reactors, gradient zones of substrate, dissolved oxygen, pH and other parameters are formed due to inefficient mixing and *E. coli* cells respond to these environmental changes by modulating their metabolism (Schweder *et al.*, 1999). For instance, *E. coli* responds to glucose excess and oxygen limitation by shifting metabolism from oxidative respiration to mixed-acid fermentation, resulting in overflow metabolism (Enfors *et al.*, 2001). Under these conditions, not only the mixed-acid fermentation products accumulate, but also pyruvate (Soini *et al.*, 2008, I). Pyruvate excess present intracellularly increases the metabolic flux going to ncBCAA biosynthesis through the sequential keto acid chain elongation from pyruvate to α -ketocaproate over α -ketobutyrate and α -ketovalerate by the actuation of the *leu* operon-encoded enzymes (Apostol *et al.*, 1997). This hypothesis is supported by the observations reported by Soini *et al.* (2011): the combination of oxygen limitation with a constant glucose supply in a two-compartment STR-PFR scale-down reactor showed a significant impact on enhancing norvaline biosynthesis due to pyruvate accumulation in a recombinant *E. coli* cultivation. Furthermore, Soini *et al.* (2008, I) originally reported accumulation of pyruvate-based amino acids such the ncBCAAs norleucine and norvaline as well as alanine and valine in a standard STR fed-batch *E. coli* cultivation under glucose excess and induced oxygen limitation upon a stirrer downshift. As mentioned before, concentration gradients happening in large industrial-scale reactors due to deficient mixing can be also simulated in scale-down reactors in the laboratory. In this investigation large-scale effects are reproduced in different fermentation systems by combining pyruvate pulsing and O₂ limitation. This novel cultivation strategy might represent more accurately the physiological behavior of bacterial cultivations taking place in large scale reactors.

Research hypothesis 2 was confirmed according to results reported in Figure 20 and Figure 28 (and additionally, Figure S71 and Figure S72). The cultivation strategy consisting of combining pyruvate pulsing and O₂ limitation was demonstrated to be an alternative method to represent large-scale effects in both shake flask and mini-reactor scales since concentrations of norleucine and norvaline increased by using the aforementioned strategy. Similar results concerning increase of ncBCAA biosynthesis under those cultivation conditions were also observed in a 15L reactor scale (Figure 32 and Figure 33). In this case, overflow metabolism triggered by pyruvate pulsing and O₂ limitation could also be demonstrated by analyzing concentrations of acetate and formate (Figure 34). As shown in Figure 20, it could also be demonstrated that the triggered ncBCAA biosynthesis is directly proportional to the amount of supplemented pyruvate.

According to the defined research hypothesis 3, the use of O₂ limitation in combination with pyruvate pulses to mimic large-scale effects may have an advantage with respect to other strategies based on glucose pulses. Pyruvate is the core substrate of the metabolic pathway leading to ncBCAA biosynthesis. Hence, when adding pyruvate to the cultivation medium, this enters the cell and accumulates intracellularly, rapidly triggering ncBCAA formation through the sequential keto acid chain elongation by the actuation of the enzymes encoded by the *leu* operon. Unlike pyruvate, when using glucose pulses, large-scale effects (i. e. increase of ncBCAA biosynthesis) can first be reported once glucose is converted to pyruvate and this starts accumulating in the cell. In this study, the intracellular soluble fraction of an *E. coli* cultivation reported the first significant increase of norleucine and norvaline concentrations around 1.5h after triggering pyruvate pulses and oxygen limitation (Figure 32). In a similar study, Soini *et al.* (2008, I) induced oxygen limitation in an *E. coli* cultivation at a high glucose concentration. In that study, the first significant increase of norvaline levels was reported around 1h after applying the aforementioned cultivation conditions, which is even earlier than reported in this study by using pyruvate pulses (1.5h). This is not in agreement with what was previously argued, being research hypothesis 3 rejected. This might be due to the fact that the employed glucose concentration in the study of Soini *et al.* (2008, I) is much higher than the pyruvate concentration used in this study, which was a total of 5 g/L. That might also be explained due to the presence in *E. coli* of different uptake/transport systems for both carbon substrates and its regulation. Kreth *et al.* (2013) suggested the existence of at least three transport systems for pyruvate in order to equilibrate intracellular pyruvate concentrations: an inducible uptake system (Usp system), a constitutive uptake system (PrvT system) and an excretion system. The first two systems were demonstrated to be controlled by catabolite repression. Hence, it might be that glucose remains present in the cultivation medium during pyruvate pulsing inhibit transport of pyruvate into the cell or that pyruvate accumulation into the cell is reverted by activation of the pyruvate excretion system. Glucose can also be uptaken by means of different transport systems: the glucose PTS (phosphotransferase system) system, the mannose PTS system, the galactose ABC transporter, the galactose permease and the maltose ABC transporter (Steinsiek and Bettenbrock, 2012).

Furthermore, the inclusion body fraction of *E. coli* cultivation showed the first significant increase of ncBCAA levels around 3.5h after triggering pyruvate pulses and oxygen limitation (Figure 33), which is 2h later than reported in the intracellular soluble protein fraction. This suggests that metabolic effects triggered by pyruvate pulsing combined with O₂ limitation occur first in the cytosol (intracellular soluble protein fraction), where ncBCAAs are synthesized. Afterwards, translation machinery would mis-incorporate those previously synthesized ncBCAA present in the cytosol into the nascent recombinant proteins (inclusion body fraction). This is also supported by the fact that both norvaline and norleucine present a higher concentration in the intracellular soluble fraction than in the inclusion bodies. Accordingly, research hypothesis 4 was confirmed.

Unlike norvaline and norleucine, concentration of β -methylnorleucine did not significantly change after exposing an *E. coli* cultivation to pyruvate pulsing and O₂ limitation with respect to the non-altered cultivation at both tested scales (Figure 32). This suggests that the metabolic effect triggered by pyruvate pulsing and O₂ limitation has a limited influence on β -methylnorleucine synthesis, in contrast to norleucine and norvaline. The fact that β -methylnorleucine is the ncBCAA, the biosynthetic pathway of which is the longest starting from pyruvate, also supports this hypothesis. Norvaline and norleucine biosynthesis require the enzymes of the *leu* operon while β -

methylnorleucine biosynthesis demands the additional actuation of the enzymes of the *ilv* operon. Observed results might as well suggest that β -methylnorleucine is synthesized by an independent alternative metabolic pathway.

In order to proof research hypothesis 5, the relative proportion of the three target ncBCAA was evaluated taking into consideration the different analysed protein fractions, cultivation conditions and tested scale. At shake flask level and mini-reactor scale, norvaline and β -methylnorleucine concentrations reported in the intracellular soluble fraction under both standard cultivation conditions and under pyruvate pulsing and O₂ limitation conditions were higher than norleucine. However, at 15 L reactor scale, norvaline and norleucine reported higher concentrations with respect to β -methylnorleucine in the intracellular soluble fraction under both tested cultivation conditions. As opposed to intracellular soluble fraction, analysis of ncBCAA in the inclusion body fraction at both mini-reactor and 15L reactor scale under both tested cultivation conditions revealed that norleucine concentration is significantly higher than norvaline and that β -methylnorleucine could not be detected. This result is interesting taking into account that the expressed recombinant protein contains only 3 methionines while 14 leucines and that norleucine and norvaline are analog to methionine and leucine, respectively. Hence, a higher mis-incorporation of norvaline would be actually expected. These observations suggest that the amount of a certain cBCAA in the amino acid sequence of a recombinant protein is not the main factor determining mis-incorporation of the respective ncBCAA analog in such recombinant protein. Accordingly, research hypothesis 5 was rejected. In addition, these results might lead to the assumption that, despite being norvaline and norleucine present in similar concentration levels intracellularly, the probability of misaminoacylation by met-tRNA is higher than by leu-tRNA. Furthermore, in both inclusion body and intracellular soluble fraction, norvaline and norleucine concentrations were higher at 15 L reactor than at mini-reactor scale. Although not comparable to industrial scale reactors, it might be that mixing in the 15 L reactor is inefficient, thus triggering formation of O₂ and glucose gradients, which, in turn, lead to ncBCAA formation, even under standard cultivation conditions.

The effect of recombinant protein induction in ncBCAA biosynthesis was also investigated. According to reported results in Figure 33, IPTG-mediated induction of recombinant protein expression is the main factor triggering ncBCAA formation under standard cultivation conditions, hence confirming research hypothesis 6. The recombinant mini-proinsulin expressed in this study consisted of a total of 96 amino acids, 14 of which were leucine residues (14.5%) while an average *E. coli* protein only contains 8.4% leucine (Neidhardt and Umbarger, 1996). Overexpression of leucine-rich recombinant proteins cause depletion of the intracellular leucine pool which, in turn, causes de-regulation of the enzymes encoded by the *leu* operon in order to counteract leucine limitation (Bogosian *et al.*, 1989; Apostol *et al.*, 1997). In parallel, activation of the *leu* operon triggers ncBCAA biosynthesis through a number of sequential chain elongation reactions starting from pyruvate (Sycheva *et al.*, 2007). Furthermore, under amino acid starvation conditions such as the aforementioned leucine depletion, transcriptional regulation of numerous genes governing the BCAA biosynthetic pathway is driven by the global RelA/SpoT modulon. Hence, operons *leuABCD*, *ilvGMEDA* and *ilvBN* as well as genes *ilvC* and *thrA* are up-regulated whereas operon *ilvIH* is down-regulated by the global (p)ppGpp regulator after amino acid starvation (Traxler *et al.*, 2008; Umbarger, 1996; Tedin and Norel, 2001; Baccigalupi *et al.*, 1995). Considering this regulatory configuration and taking into account that *ilvG* is not active in the *E. coli* strain used in this study (*E. coli* K-12 BW25113), metabolic flux would be redirected to leucine, valine and ncBCAA biosynthesis.

6.6 Screening of *geneX*-tunable *E. coli* strains

Research hypothesis 1 was defined by taking into consideration the BCAA metabolic pathway. The following strategies were pursued by modulating regulation of target genes in *E. coli* in order to reduce ncBCAA biosynthesis: (i) limit conversion of pyruvate to α -ketobutyrate, (ii) limit transformation of threonine to α -ketobutyrate and (iii) limit conversion of α -ketobutyrate to α -ketovalerate. Strategy (i) might be achieved by down-regulating operon *leuABCD* but also by up-regulating operon *ilvBN*, strategy (ii) may be realised by down-regulating the *thr* genes as well as *ilvA*. Strategy (iii) could be accomplished by down-regulating operon *leuABCD* and up-regulating *ilvIH*, *ilvGM* and *ilvC*. According to this hypothesis novel *E. coli* strain mutants were genetically engineered so that the expression of single target genes (*leuA*, *thrA*, *ilvA*, *ilvC*, *ilvIH*, *ilvBN* and *ilvGM*) could be modulated in order to evaluate the effect of genetic modulation in ncBCAA biosynthesis. Screening of the engineered *E. coli* mutants was performed in a 10 mL mini-reactor under standard cultivation conditions. Additional screening of mutants under cultivation conditions triggering ncBCAA, i. e. pyruvate pulses and O₂ limitation, allowed elucidating performance of mentioned *E. coli* mutants under large-scale cultivation conditions.

In order to support discussion of the reported results following hypothesis already described in literature were assumed: (i) since the recombinant protein expressed in this study has a high leucine-content, a depletion of the intracellular leucine pool after IPTG induction is expected, thus causing de-regulation of the *leu* operon, which, in turn, triggers a relative increase of ncBCAA biosynthesis (Bogosian *et al.*, 1989; Apostol *et al.*, 1997); (ii) under cultivation conditions based on pyruvate pulses and O₂ limitation an intracellular accumulation of pyruvate and, consequently, an increase of the metabolic flux through leucine, valine and ncBCAA pathways is expected (Soini *et al.*, 2008, I). In addition, for each scenario, transcriptional and post-translational regulation of target genes (summarized in Table 1, section 2.2.5) was also taken into consideration for discussion.

Down-regulation of *leuA* was hypothesized to limit ncBCAA biosynthesis by restricting the successive keto acid chain elongation reactions starting from pyruvate until ncBCAA formation. Some investigations previously demonstrated that knocking-out one or more *leu* genes reduces ncBCAA biosynthesis (Fenton *et al.*, 1994; Bogosian *et al.*, 1989). Results obtained for *leuA*-tunable *E. coli* under standard cultivation conditions in this study are in accordance with the logic of the metabolic pathway since, for the mentioned *E. coli* mutant, norvaline and norleucine concentrations present in both tested protein fractions progressively increased by adding increasing concentrations of L-arabinose into the medium, i.e. by increasing *leuA* expression. Interestingly, under cultivation conditions subjected to pyruvate pulses and O₂ limitation, increasing *leuA* expression did not translate into significant variation of ncBCAA concentrations but those remained always higher than under standard cultivation conditions. This observation might be explained due to a saturation of the metabolic pathway comprising the consecutive keto acid chain elongation reactions starting from pyruvate and leading to ncBCAA caused by the expected intracellular accumulation of pyruvate and de-regulation of *leu* under such cultivation conditions, so that higher *leuA* expression would then not trigger more carbon flux to ncBCAA formation.

Up-regulation of *ilvC* was hypothesized to trigger reduction of ncBCAA biosynthesis, since that would stimulate metabolic flux from α -ketobutyrate through the isoleucine biosynthetic pathway, thereby relatively reducing α -ketobutyrate disponibility for ncBCAA formation. Results confirmed hypothesis

since, for *ilvC*-tunable *E. coli* under standard cultivation conditions, norvaline and norleucine concentrations present in both tested protein fractions progressively decreased by adding increasing concentrations of L-arabinose into the medium, i.e. by increasing *ilvC* expression. The same trend was observed when pyruvate pulsing and O₂ limitation conditions were applied. As expected, under those cultivation conditions, ncBCAA concentrations were higher.

AHAS III has more substrate preference for α -ketobutyrate than pyruvate (Barak *et al.*, 1987; Salmon *et al.*, 2006; Vinogradow *et al.*, 2006). Hence, an increase in AHAS III concentration would favor the metabolic flux from α -ketobutyrate into α -acetohydroxy-butyrate at the expense of the alternative enzymatic reaction leading to ncBCAA synthesis. Results confirmed this hypothesis since for *ilvIH*-tunable *E. coli* under standard cultivation conditions, norvaline and norleucine concentrations present in both tested protein fractions progressively decreased by adding increasing concentrations of L-arabinose into the medium, i. e. by increasing *ilvIH* expression. Surprisingly, under cultivation conditions subjected to pyruvate pulses and O₂ limitation, the opposite trend was observed: increasing *ilvIH* expression was translated into an increase of ncBCAA concentrations. The simultaneous presence of multiple factors playing a role in the regulation of the BCAA metabolic pathway under mentioned cultivation conditions made challenging to find a plausible explanation for those observations.

The *ilvGM*-tunable *E. coli* strain also reported a reduction of norvaline and norleucine concentrations compared with the wild type *E. coli* strain under both tested cultivation conditions. As expected, under cultivation conditions subjected to pyruvate pulsing and O₂ limitation, ncBCAA concentrations were higher. Moreover, the highest ncBCAA reduction observed among all tested mutant strains was reported for this mutant. This might be because AHAS II shows the highest substrate preference for α -ketobutyrate among the AHAS isozymes, being then most of the metabolic flux from α -ketobutyrate directed to the isoleucine pathway. In addition, k_{cat}/K_m of AHAS II is about 20-fold higher than AHAS III (Barak *et al.*, 1987; Salmon *et al.*, 2006; Vinogradow *et al.*, 2006). However, increasing of *ilvGM* expression did not translate into significant variation of ncBCAA concentrations. That might be explained by a negative feedback regulation due to an excess of free isoleucine, which would decrease AHAS II activity (Salmon *et al.*, 2006), thus counteracting the increased AHAS II activity driven by L-arabinose addition. Thus, in the ideal case, *ilvGM* expression should be up-regulated as much as possible but without leading to an accumulation of isoleucine.

Unexpectedly, the *ilvBN*-tunable *E. coli* strain reported the opposite behavior than the tested *ilvIH*- and *ilvGM*-tunable *E. coli* strains under standard cultivation conditions: norvaline and norleucine concentrations progressively increased by adding increasing concentrations of L-arabinose into the medium, i. e. by increasing *ilvBN* expression. As opposed to AHAS II and III, AHAS I prefers pyruvate than α -ketobutyrate as substrate (Barak *et al.*, 1987; Salmon *et al.*, 2006; Vinogradow *et al.*, 2006). Hence, an increase in AHAS I concentration would favor the metabolic flux from pyruvate into the valine and leucine biosynthetic pathway at the expense of pyruvate transformation to α -ketobutyrate by the *leu* operon. However, up-regulation of *ilvBN* might trigger overproduction of valine, which is demonstrated to inhibit enzymatic activity of AHAS III by feedback regulation and to activate enzymatic activity of L-threonine dehydratase (*ilvA*-encoded enzyme) (Salmon *et al.*, 2006). According to this configuration, more α -ketobutyrate would be produced from the threonine pathway and, since activity of AHAS III is feedback regulated, α -ketobutyrate might then preferably enter the ncBCAA biosynthetic pathway to the detriment of the isoleucine biosynthetic pathway. This

metabolic configuration might explain why, as opposed to first hypothesized, a down-regulation of operon *ilvBN* triggers reduction of ncBCAA biosynthesis. The same trend was observed when pyruvate pulsing and O₂ limitation conditions were applied but, interestingly, ncBCAA concentrations were lower than under standard conditions. The simultaneous presence of multiple factors playing a role in the regulation of the BCAA metabolic pathway under mentioned cultivation conditions made challenging to find a plausible explanation for those observations.

The *thrA*-tunable *E. coli* strain showed a reduction of norvaline and norleucine concentrations in both tested protein fractions compared with the wild type strain under both tested cultivation conditions. As expected, under cultivation conditions subjected to pyruvate pulsing and O₂ limitation, ncBCAA concentrations were higher. However, effect of increasing L-arabinose concentrations did not show a clear effect on ncBCAA concentrations. With 0.4% L-ara, a significant decrease in norvaline and norleucine was observed. However, a further increase in *thrA* expression when using higher L-ara concentrations did not show any further variation of ncBCAA concentration if compared with 0.4% L-ara. That might be explained by a bottleneck taking place in the metabolic pathway downstream of *thrA*, so that a higher *thrA* expression does not translate in higher α -ketobutyrate production. In addition, it is well known that L-threonine allosterically inhibits enzyme activity of *thrA*-encoded enzyme so that an excessive L-ara induction triggers accumulation of threonine, then causing feedback inhibition of enzymatic activity, hence annulling the higher expected *thrA* expression. This observation might indicate that the threonine metabolic pathway is not the one redirecting more flux to α -ketobutyrate but the pyruvate pathway.

The *ilvA*-tunable *E. coli* strain showed that, mainly in the intracellular protein soluble protein fraction, norvaline and norleucine concentrations report a reduction compared with control strain under both tested cultivation conditions but effect of increasing L-arabinose concentrations did not show a clear trend on variation of ncBCAA concentrations. As expected, under cultivation conditions subjected to pyruvate pulsing and O₂ limitation, ncBCAA concentrations were higher. With 0.05% L-ara, a significant decrease in norvaline and norleucine levels was observed. However, a further increase in *ilvA* expression when using higher L-ara concentrations did not show any further reduction of ncBCAA concentration if compared with 0.05% L-ara. As for *thrA*, that might be explained by a bottleneck in the metabolic pathway downstream of *ilvA*. Moreover, it is well known that isoleucine allosterically inhibits enzyme activity of *ilvA*-encoded enzyme. May then be that an excessive L-ara induction triggers accumulation of isoleucine, then causing feedback inhibition of enzymatic activity, hence annulling the higher expected *ilvA* expression. As stated before, this observation might indicate that the threonine metabolic pathway is not the one redirecting more flux to α -ketobutyrate but the pyruvate pathway.

Unlike norvaline and norleucine, concentration of β -methylnorleucine did not significantly change after modulating expression of target genes. This suggests that the metabolic effect triggered by tuning expression of target genes involved in the BCAA biosynthetic pathway has a limited influence on β -methylnorleucine synthesis, in contrast to norleucine and norvaline. As aforementioned, the fact that β -methylnorleucine is the ncBCAA, the biosynthetic pathway of which is the longest starting from pyruvate, also supports this hypothesis. Norvaline and norleucine biosynthesis require the enzymes of the *leu* operon to transform pyruvate to the ncBCAA precursors α -ketovalerate and α -ketocaproate while β -methylnorleucine biosynthesis demands the additional actuation of the enzymes of the *ilv* operon in order to convert α -ketovalerate into the β -methylnorleucine precursor α -keto- β -methylcaproate. Observed results might as well suggest that β -methylnorleucine is

synthesized by an independent alternative metabolic pathway. However, no references were found in the literature supporting this hypothesis.

According to the previous discussion, research hypothesis 1 was partially accepted.

7. Conclusions and Outlook

This thesis was aimed to close the scientific gap by developing a new scientific approach in order to reduce ncBCAA biosynthesis and subsequent mis-incorporation into recombinant proteins since all the alternative strategies published so far present numerous disadvantages which challenge their effective application. In this study we demonstrated that ncBCAA biosynthesis can be exogenously controlled by fine tuning expression of target genes involved in the BCAA biosynthetic pathway. By screening different engineered tunable *E. coli* mutants in a mini-reactor system we demonstrated that an up-regulation of *ilvC*, *ilvIH* and *ilvGM* and down-regulation of *leuA* and *ilvBN* trigger a reduction of norvaline and norleucine biosynthesis and mis-incorporation into recombinant mini-proinsulin expressed in this study. Concerning target genes *ilvA* and *thrA*, results suggest that the threonine pathway is not the one redirecting more metabolic flux to α -ketobutyrate. Among the tested genes, up-regulation of *ilvIH* and *ilvGM* showed the highest reduction of ncBCAA biosynthesis and mis-incorporation. Strain screening was performed in fed-batch mode under standard cultivation conditions as well as under cultivation conditions subjected to pyruvate pulsing and O₂ limitation. The latter cultivation strategy was demonstrated to mimic large-scale effects since biosynthesis of ncBCAA and metabolites of the overflow metabolism were reported under those conditions. This cultivation approach represents a novel scale-down strategy which could contribute to accelerate the process of strain screening. Interestingly, norleucine was the most mis-incorporated ncBCAA and β -methylnorleucine levels did not significantly change under tested experimental conditions, which may suggest that β -methylnorleucine is synthesized by an alternative unknown metabolic pathway. Potential *ilvIH* and *ilvGM*-tunable *E. coli* strains showing a preferred protein impurity profile during screening in the mini-reactor system were further verified in a 15L reactor under cultivation conditions subjected to pyruvate pulsing and O₂ limitation, reporting same recombinant mini-proinsulin production and a highly significant reduction of ncBCAA mis-incorporation into the recombinant protein in comparison with the non-engineered *E. coli* strain. This is in accordance with what was reported at mini-reactor level, thus confirming the reliability and robustness of the aforementioned strains. These novel *E. coli* strains might then be employed as expression hosts in large-scale reactors for industrial production of recombinant proteins with a reduced ncBCAA mis-incorporation profile. Testing of the engineered *E. coli* strains in alternative scale-down reactors such as a 2-compartment scale-down reactor (Limberg *et al.*, 2016) would give more hints about their applicability in industrial scale recombinant protein production processes.

In order to ensure optimal expression of the target gene it was necessary to expose the corresponding engineered tunable *E. coli* strain to a certain L-arabinose concentration and this was only possible by exogenously adding the inducer molecule into the system, which is not optimal from an economic and operational point of view if recombinant protein production is intended in industrial scale. The ideal scenario would be that the optimal expression level of a target gene resulting in a reduced ncBCAA mis-incorporation would be maintained constant during the *E. coli* fermentation process without the need of adding external chemical compounds. A number of genetic engineering strategies such as promoter engineering and antisense RNA technology could be applied in order to achieve that preferred scenario. The development of novel endogenous constitutive promoters through rational design or random mutagenesis allowing optimal expression levels of a target gene would allow production of recombinant proteins with a reduced ncBCAA mis-incorporation profile without the need of transforming additional plasmids into the *E. coli* expression

host. This approach might be appropriate for both genetic up- and down-regulation purposes. An antisense RNA (asRNA) is a single stranded RNA complementary to a certain mRNA so that both RNA molecules can hybridize, blocking mRNA translation. This strategy might be useful for genetic down-regulation purposes. However, additional transformation of a plasmid enabling asRNA expression might be necessary in this case.

Moreover, optimal expression levels of single target genes leading to a reduction of ncBCAA mis-incorporation into recombinant mini-proinsulin were investigated in this current thesis. However, it might be also very interesting to investigate expression regulation of two or more target genes simultaneously in order to achieve further reduction of ncBCAA mis-incorporation that might not be possible by just considering a single gene. This might be possible by using alternative tunable promoters, one different for each target gene. For instance, in order to evaluate effect of expression regulation of two different target genes simultaneously in an *E. coli* host, an arabinose-inducible promoter could be used for one gene while a XylSPm promoter could be employed for the other. Taking into consideration results obtained in this investigation, the following two genetic combinations are predicted to report the most improved recombinant protein impurity profile: (i) down-regulation of gene *leuA* and repair or up-regulation of operon *ilvGM*, (ii) down-regulation of gene *leuA* and up-regulation of operon *ilvIH*. The use of alternative molecular methods such as promoter engineering and antisense RNA technology might also be considered for this purpose.

In order to better understand how endogenous genetic regulation affects ncBCAA biosynthesis it would be interesting to analyze the content of free amino acids present in the cytosol. In this study the amino acid content was analyzed in the inclusion body fraction and in the intracellular protein soluble fraction (including the free amino acids). Since amino acid depletion or amino acid excess triggers several genetic regulation mechanisms affecting expression of genes involved in the BCAA biosynthetic pathway at both transcriptional (e. g. attenuation, stringent response) and posttranslational level (e. g. feedback inhibition) that could modulate ncBCAA biosynthesis, such additional data would be valuable to complete the discussion of some of the results generated in this thesis. Furthermore, the experimental design and analysis carried out in this thesis were considered sufficient to determine if an up- or down-regulation of a certain target gene causes a reduction of ncBCAA biosynthesis, by measuring ncBCAA levels under different inducer concentrations. However, it would be interesting to analyze the actual expression levels of each tested target gene under different L-arabinose concentrations by means of transcriptomics or proteomics technologies in order to identify the optimal expression level of a target gene triggering the most reduction of ncBCAA biosynthesis. In a most preferred scenario, the availability of metabolic flux data would contribute to obtain a clear picture of the metabolic stand of the BCAA biosynthetic pathway resulting from each experimental scenario.

Further strategies aiming reduction of ncBCAA mis-incorporation, which are not yet available in the state of the art and that, together with the approach investigated in this thesis, might as well be considered as potential comprise: (i) substitution of codons reporting higher ncBCAA mis-incorporation from the coding region of a gene of interest, (ii) protein engineering to increase specificity of biosynthetic enzymes for certain α -keto acids, (iii) protein engineering to increase specificity of aminoacyl tRNA synthetases for the canonical amino acids, and (iv) repair of gene *ilvG* in the genome of *E. coli* K-12 strains to recover AHAS II activity.

8. References

- Abu-Absi, N., Inlow, D., Macdonald, A. C., Poulhzan, M. (2008). Preventing norvaline and norleucine mis-incorporation in recombinant proteins. US Patent Application WO 2007/103521 A3.
- Afroz, T., Biliouris, K., Kaznessis, Y., & Beisel, C. L. (2014). Bacterial sugar utilization gives rise to distinct single-cell behaviours. *Molecular microbiology*, 93(6), 1093-1103.
- Ahmadi, M. K., Fang, L., Moscatello, N., & Pfeifer, B. A. (2016). E. coli metabolic engineering for gram scale production of a plant-based anti-inflammatory agent. *Metabolic Engineering*, 38, 382-388.
- Alba, B. M. & Gross, C. A. (2004). Regulation of the *Escherichia coli* sigma-dependent envelope stress response. *Mol Microbiol* 52, 613–619.
- Alvarez-Carreño, Claudia, Arturo Becerra, and Antonio Lazcano. "Norvaline and Norleucine may have been more abundant protein components during early stages of cell evolution." *Origins of Life and Evolution of Biospheres* 43.4-5 (2013): 363-375.
- Amann, E., Brosius, J., & Ptashne, M. (1983). Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene*, 25(2), 167-178.
- Anane E, Lòpez CDC, Neubauer P, Cruz Bournazou MN (2017) Modelling overflow metabolism in *Escherichia coli* by acetate cycling. *Biochem Engin J* 125(15):23–30
- Anane, E., Sawatzki, A., Neubauer, P., & Cruz-Bournazou, M. N. (2019). Modelling concentration gradients in fed-batch cultivations of *E. coli*—towards the flexible design of scale-down experiments. *Journal of Chemical Technology & Biotechnology*, 94(2), 516-526.
- Andersen DC, Swartz J; Ryll T, Lin N, Snedecor B. 2001. Metabolic Oscillations in an *E. coli* Fermentation. *Biotechnol. Bioen.* 75(2): 212-218.
- Apostol, I., Levine, J., Lippincott, J., Leach, J., Hess, E., Glascock, C. B., ... & Blackmore, R. (1997). Incorporation of norvaline at leucine positions in recombinant human hemoglobin expressed in *Escherichia coli*. *Journal of Biological Chemistry*, 272(46), 28980-28988.
- Arditti, R., Grodzicker, T., & Beckwith, J. (1973). Cyclic adenosine monophosphate-independent mutants of the lactose operon of *Escherichia coli*. *Journal of bacteriology*, 114(2), 652-655.
- Arfin, S.M., Ratzkin, B. & Umbarger, H.E. (1969) The metabolism of valine and isoleucine in *Escherichia coli*. XVII. The role of induction in the derepression of acetohydroxy acid isomeroreductase. *Biochem Biophys Res Commun* 37: 902–908.
- Aris, A., Corchero, J. L., Benito, A., Carbonell, X., Viaplana, E., & Villaverde, A. (1998). The expression of recombinant genes from bacteriophage lambda strong promoters triggers the SOS response in *Escherichia coli*. *Biotechnology and bioengineering*, 60(5), 551-559.
- Artsimovitch, I., Patlan, V., Sekine, S. I., Vassilyeva, M. N., Hosaka, T., Ochi, K., ... & Vassilyev, D. G. (2004). Structural basis for transcription regulation by alarmone ppGpp. *Cell*, 117(3), 299-310.

- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2(1).
- Baccigalupi, L., Marasco, R., Ricca, E., De Felice, M., & Sacco, M. (1995). Control of *ilvH* transcription during amino acid downshift in stringent and relaxed strains of *Escherichia coli*. *FEMS microbiology letters*, 131(1), 95-98.
- Bailey, J. E. (1991, June). Toward a science of metabolic engineering. *American Association for the Advancement of Science*.
- Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T., & Valla, S. (2013). A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. *Microbial cell factories*, 12(1), 1.
- Barak, Z., Chipman, D. M., & Gollop, N. A. T. A. N. (1987). Physiological implications of the specificity of acetohydroxy acid synthase isozymes of enteric bacteria. *Journal of bacteriology*, 169(8), 3750-3756;
- Basturea, G. N., Zundel, M. A., & Deutscher, M. P. (2011). Degradation of ribosomal RNA during starvation: comparison to quality control during steady-state growth and a role for RNase PH. *Rna*, 17(2), 338-345.
- Beals JM, DeFelippis MR, Kovach PM. Chapter 12: Insulin. In: Crommelin DJA, Sindelar RD, Meibohm B (eds): *Pharmaceutical biotechnology: fundamentals and application*. 3rd ed. 12:265-280, Informa Healthcare (2008).
- Biermann, M., Linnemann, J., Knüpfer, U., Vollstädt, S., Bardl, B., Seidel, G., & Horn, U. (2013). Trace element associated reduction of norleucine and norvaline accumulation during oxygen limitation in a recombinant *Escherichia coli* fermentation. *Microbial cell factories*, 12(1), 1.
- Binder, D., Bier, C., Grünberger, A., Drobiez, D., Hage-Hülsmann, J., Wandrey, G., ... & Jaeger, K. E. (2016). Photocaged Arabinose: A Novel Optogenetic Switch for Rapid and Gradual Control of Microbial Gene Expression. *ChemBioChem*, 17(4), 296-299.
- Binder, D., Grünberger, A., Loeschcke, A., Probst, C., Bier, C., Pietruszka, J., ... & Drepper, T. (2014). Light-responsive control of bacterial gene expression: precise triggering of the *lac* promoter activity using photocaged IPTG. *Integrative biology*, 6(8), 755-765.
- Biryukova, I. V., Krylov, A. A., Kiseleva, E. M., Minaeva, N. I., & Mashko, S. V. (2010). Construction of the new *Escherichia coli* K12 MG 1655 novel strain with improved growth characteristics for application in metabolic engineering. *Russian journal of genetics*, 46(3), 308-314.
- Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Karunakaran, P., & Valla, S. (1997). Improved broad-host-range RK2 vectors useful for high and low regulated gene expression levels in gram-negative bacteria. *Plasmid*, 38(1), 35-51.
- Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ... & Gregor, J. (1997). The complete genome sequence of *Escherichia coli* K12. *Science*, 277(5331), 1453-1462.
- Bloor, A. E., & Cranenburgh, R. M. (2006). An efficient method of selectable marker gene excision by Xer recombination for gene replacement in bacterial chromosomes. *Applied and environmental microbiology*, 72(4), 2520-2525.

- Bogosian, G., O'Neil, J. P., Smith, H. Q. (2013). Prevention of incorporation of non-standard amino acids into proteins. US Patent 8603781.
- Bogosian, G., Violand, B. N., Dorward-King, E. J., Workman, W. E., Jung, P. E., & Kane, J. F. (1989). Biosynthesis and incorporation into protein of norleucine by *Escherichia coli*. *Journal of Biological Chemistry*, 264(1), 531-539.
- Bowden, K. E., Wiese, N. S., Perwez, T., Mohanty, B. K., & Kushner, S. R. (2017). The rph-1 encoded truncated RNase PH protein inhibits RNase P maturation of pre-tRNAs with short leader sequences in the absence of RppH. *Journal of bacteriology*, JB-00301.
- Brautaset, T., Lale, R., & Valla, S. (2009). Positively regulated bacterial expression systems. *Microbial biotechnology*, 2(1), 15-30.
- Brautaset, T., Petersen, S. B., & Valla, S. (2000). In vitro determined kinetic properties of mutant phosphoglucomutases and their effects on sugar catabolism in *Escherichia coli*. *Metabolic engineering*, 2(2), 104-114.
- Brown, T. D. K., Jones-Mortimer, M. C., & Kornberg, H. L. (1977). The enzymic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. *Microbiology*, 102(2), 327-336.
- Brückner, R., & Titgemeyer, F. (2002). Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS microbiology letters*, 209(2), 141-148.
- Brunner et al. (1997). Fermentation media and methods for controlling norleucine in polypeptides. US Pat. No. 5698418
- Busby, S., & Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *Journal of molecular biology*, 293(2), 199-213.
- 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 Engineering*, 18(3), 171-180.
- Calos, M. P. (1978). DNA sequence for a low-level promoter of the lac repressor gene and anup'promoter mutation. *Nature*, 274, 762-765.
- Calos, M. P., & Miller, J. H. (1980). DNA sequence alteration resulting from a mutation impairing promoter function in the lac repressor gene. *Molecular and General Genetics MGG*, 178(1), 225-227
- Calos, M. P., & Miller, J. H. (1981). The DNA sequence change resulting from the I Q1mutation, which greatly increases promoter strength. *Molecular and General Genetics MGG*, 183(3), 559-560.
- Campos-Bermudez, V. A., Bologna, F. P., Andreo, C. S., & Drincovich, M. F. (2010). Functional dissection of *Escherichia coli* phosphotransacetylase structural domains and analysis of key compounds involved in activity regulation. *The FEBS journal*, 277(8), 1957-1966.
- Cardona, S. T., & Valvano, M. A. (2005). An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid*, 54(3), 219-228.
- Carneiro, S., Ferreira, E. C., & Rocha, I. (2013). Metabolic responses to recombinant bioprocesses in *Escherichia coli*. *Journal of biotechnology*, 164(3), 396-408

- Carrier, T. A., & Keasling, J. D. (1999). Investigating autocatalytic gene expression systems through mechanistic modeling. *Journal of theoretical biology*, 201(1), 25-36.
- Chen, C. F., Lan, J., Korovine, M., Shao, Z. Q., Tao, L., Zhang, J., & Newman, E. B. (1997). Metabolic regulation of *Irp* gene expression in *Escherichia coli* K12. *Microbiology*, 143(6), 2079-2084.
- Chen, S., Iannolo, M., & Calvo, J. M. (2005). Cooperative binding of the leucine-responsive regulatory protein (*Lrp*) to DNA. *Journal of molecular biology*, 345(2), 251-264.
- Chen, X., Ma, J. J., Tan, M., Yao, P., Hu, Q. H., Eriani, G., & Wang, E. D. (2011). Modular pathways for editing non-cognate amino acids by human cytoplasmic leucyl-tRNA synthetase. *Nucleic acids research*, 39(1), 235-247.
- Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: property, design and functionality. *Advanced drug delivery reviews*, 65(10), 1357-1369.
- Chen, Z., Sun, X., Li, Y., Yan, Y., & Yuan, Q. (2017). Metabolic engineering of *Escherichia coli* for microbial synthesis of monolignols. *Metabolic Engineering*.
- Choi, J. H., Keum, K. C., & Lee, S. Y. (2006). Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chemical Engineering Science*, 61(3), 876-885.
- Cirino, P. C., Tang, Y., Takahashi, K., Tirrell, D. A., & Arnold, F. H. (2003). Global incorporation of norleucine in place of methionine in cytochrome P 450 BM-3 heme domain increases peroxygenase activity. *Biotechnology and bioengineering*, 83(6), 729-734.
- Cohen, G. N., & Munier, R. (1959). Effets des analogues structuraux d'aminioacides sur la croissance, la synthese de proteines et la synthese d'enzymes chez *Escherichia coli*. *Biochem. Biophys. Acta* 31(2): 347-356).
- Cohen, S. N., Chang, A. C., & Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proceedings of the National Academy of Sciences*, 69(8), 2110-2114.
- Court, D. L., Sawitzke, J. A., & Thomason, L. C. (2002). Genetic engineering using homologous recombination. *Annual review of genetics*, 36(1), 361-388.
- Couturier, E., & Rocha, E. P. (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular microbiology*, 59(5), 1506-1518.
- Cowie, D. B., Cohen, G. N., Bolton, E. T., & De Robichon-Szulmajster, H. (1959). Amino acid analog incorporation into bacterial proteins. *Biochimica et biophysica acta*, 34, 39-46.
- Cvak, L., Jegorov, A., Sedmera, P., Císařová, I., Čejka, J., Kratochvíl, B., & Pakhomova, S. (2005). Norleucine, a natural occurrence in a novel ergot alkaloid γ -ergokryptinine. *Amino acids*, 29(2), 145-150.
- Cvetešić, N., Akmačić, I., & Gruić-Sovulj, I. (2013). Lack of discrimination against non-proteinogenic amino acid norvaline by elongation factor Tu from *Escherichia coli*. *Croatica Chemica Acta*, 86(1), 73-82.
- Cvetesic, N., Palencia, A., Halasz, I., Cusack, S., & Gruic-Sovulj, I. (2014). The physiological target for *LeuRS* translational quality control is norvaline. *The EMBO journal*, e201488199.

- Cvetesic, N., Perona, J. J., & Gruic-Sovulj, I. (2012). Kinetic partitioning between synthetic and editing pathways in class I aminoacyl-tRNA synthetases occurs at both pre-transfer and post-transfer hydrolytic steps. *Journal of Biological Chemistry*, 287(30), 25381-25394.
- Cvetesic, N., Semanjski, M., Soufi, B., Krug, K., Gruic-Sovulj, I., & Macek, B. (2016). Proteome-wide measurement of non-canonical bacterial mistranslation by quantitative mass spectrometry of protein modifications. *Scientific reports*, 6, 28631.
- Daber, R., Stayrook, S., Rosenberg, A., & Lewis, M. (2007). Structural analysis of lac repressor bound to allosteric effectors. *Journal of molecular biology*, 370(4), 609-619.
- Dailey, F. E., & Cronan, J. E. (1986). Acetohydroxy acid synthase I, a required enzyme for isoleucine and valine biosynthesis in *Escherichia coli* K12 during growth on acetate as the sole carbon source. *Journal of bacteriology*, 165(2), 453-460.
- Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12), 6640-6645.
- Davidson, I. (1997). Hydrolysis of samples for amino acid analysis. In *Protein Sequencing Protocols* (pp. 119-129). Humana Press, Totowa, NJ.
- De Boer, H. A., Comstock, L. J., & Vasser, M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences*, 80(1), 21-25.
- de los Rios, S., & Perona, J. J. (2007). Structure of the *Escherichia coli* leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *Journal of molecular biology*, 366(5), 1589-1602.
- DeFelice, M., and M. Levinthal. 1977. The acetohydroxy acid synthase III isoenzyme of *Escherichia coli* K12: regulation of synthesis by leucine. *Biochem. Biophys. Res. Commun.* 79:82–87
- Dong, H., Nilsson, L., & Kurland, C. G. (1996). Co-variation of trna abundance and codon usage in *Escherichia coli* at different growth rates. *Journal of molecular biology*, 260(5), 649-663.
- Dong, X., Chen, X., Qian, Y., Wang, Y., Wang, L., Qiao, W., & Liu, L. (2016). Metabolic engineering of *Escherichia coli* W3110 to produce L-malate. *Biotechnology and Bioengineering*.
- Dörschung, P.; Habermann, P.; Seipke, G.; Uhlmann, E. (1989). Mini-Proinsulin, seine Herstellung und Verwendung. Patent number: EP0347781B1.
- Egan, S. M., & Schleif, R. F. (1993). A regulatory cascade in the induction of rhaBAD. *Journal of molecular biology*, 234(1), 87-98.
- Eiteman, M. A., & Altman, E. (2006). Overcoming acetate in *Escherichia coli* recombinant protein fermentations. *Trends in biotechnology*, 24(11), 530-536.
- Enfors, S. O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., ... & Noisommit-Rizzi, N. (2001). Physiological responses to mixing in large-scale bioreactors. *Journal of biotechnology*, 85(2), 175-185
- Fahnert, B., Lilie, H., & Neubauer, P. (2004). Inclusion bodies: formation and utilisation. In *Physiological Stress Responses in Bioprocesses* (pp. 93-142). Springer, Berlin, Heidelberg.

- Fang, M., & Bauer, C. E. (2018). Regulation of stringent factor by branched-chain amino acids. *Proceedings of the National Academy of Sciences*, 201803220.
- Fenton, D., Lai, P. H., Lu, H., Mann, M., Tsai, L. (1994). Control of norleucine incorporation into recombinant proteins. US Patent 5599690.
- Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E., & Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microbial cell factories*, 8(1), 1.
- Fiedler, S., & Wirth, R. (1988). Transformation of bacteria with plasmid DNA by electroporation. *Analytical biochemistry*, 170(1), 38-44.
- Freundlich, M. (1977). Cyclic AMP can replace the relA-dependent requirement for derepression of acetohydroxy acid synthase in *E. coli* K12. *Cell*, 12(4), 1121-1126.
- Freundlich, M., Ramani, N., Mathew, E., Sirko, A., & Tsui, P. (1992). The role of integration host factor in gene expression in *Escherichia coli*. *Molecular microbiology*, 6(18), 2557-2563.
- Friden, P., Newman, T., & Freundlich, M. (1982). Nucleotide sequence of the *ilvB* promoter-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP. *Proceedings of the National Academy of Sciences*, 79(20), 6156-6160.
- Friden, P., Tsui, P., Okamoto, K., & Freundlich, M. (1984). Interaction of cyclic AMP receptor protein with the *ilvB* biosynthetic operon in *E. coli*. *Nucleic acids research*, 12(21), 8145-8160. (I)
- Friden, P., Voelkel, K., Sternglanz, R., & Freundlich, M. (1984). Reduced expression of the isoleucine and valine enzymes in integration host factor mutants of *Escherichia coli*. *Journal of molecular biology*, 172(4), 573-579. (II)
- Friedman, D. I. (1988). Integration host factor: a protein for all reasons. *Cell*, 55(4), 545-554.
- García-Fruitós, E. (2010). Inclusion bodies: a new concept. *Microbial cell factories*, 9(1), 80.
- Giacalone. (2006). Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system (vol 40, pg 355, 2006). *Biotechniques*, 40(5), 596-596.
- Gill, R. T., Valdes, J. J., & Bentley, W. E. (2000). A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metabolic engineering*, 2(3), 178-189.
- Gilles, A. M., Marliere, P., Rose, T., Sarfati, R., Longin, R., Meier, A., ... & Barzu, O. (1988). Conservative replacement of methionine by norleucine in *Escherichia coli* adenylate kinase. *Journal of Biological Chemistry*, 263(17), 8204-8209.
- Glascock, C. B., & Weickert, M. J. (1998). Using chromosomal *lacI*Q1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene*, 223(1), 221-231.
- Glick, B. R. (1995). Metabolic load and heterologous gene expression. *Biotechnology advances*, 13(2), 247-261.
- Gollop, N., Damri, B., Barak, Z. E., & Chipman, D. M. (1989). Kinetics and mechanism of acetohydroxy acid synthase isozyme III from *Escherichia coli*. *Biochemistry*, 28(15), 6310-6317.
- Görke, B., & Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology*, 6(8), 613-624.

- Grenier, F., Matteau, D., Baby, V., & Rodrigue, S. (2014). Complete genome sequence of *Escherichia coli* BW25113. *Genome announcements*, 2(5), e01038-14.
- Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. O. N. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology*, 177(14), 4121-4130.
- Habermann, P., & Wengermayer, F. (1996). Fusion protein comprising an Interleukin-2 fragment ballast portion. Patent number: 5496924.
- Habermann, Paul (1986). Fusionsproteine mit eukaryotischem Ballastanteil. Patent number: EP0229998A2.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*, 166(4), 557-580.
- Haqqi, T., Zhao, X., Panciu, A., & Yadav, S. (2002). Sequencing in the presence of betaine: Improvement in sequencing of the localized repeat sequence regions. *Journal of Biomolecular Techniques: JBT*, 13(4), 265.
- Harris, R. P., & Kilby, P. M. (2014). Amino acid mis-incorporation in recombinant biopharmaceutical products. *Current opinion in biotechnology*, 30, 45-50.
- Hartley, D. L., & Kane, J. F. (1988). Properties of inclusion bodies from recombinant *Escherichia coli*.
- Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T., & Gerdes, K. (2015). Recent functional insights into the role of (p) ppGpp in bacterial physiology. *Nature Reviews Microbiology*, 13(5), 298.
- Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., ... & Horiuchi, T. (2006). Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Molecular systems biology*, 2(1).
- Hermesen, R., Okano, H., You, C., Werner, N., & Hwa, T. (2015). A growth-rate composition formula for the growth of *E. coli* on co-utilized carbon substrates. *Molecular systems biology*, 11(4), 801.
- Hirokawa, Y., Kawano, H., Tanaka-Masuda, K., Nakamura, N., Nakagawa, A., Ito, M., ... & Ogasawara, N. (2013). Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *Journal of bioscience and bioengineering*, 116(1), 52-58.
- Huang, Y. T., Lyu, S. Y., Chuang, P. H., Hsu, N. S., Li, Y. S., Chan, H. C., ... & Li, T. L. (2009). In vitro Characterization of Enzymes Involved in the Synthesis of Nonproteinogenic Residue (2S, 3S)- β -Methylphenylalanine in Glycopeptide Antibiotic Mannopectimycin. *Chembiochem*, 10(15), 2480-2487.
- Hunter, M. F., & Parker, E. J. (2014). Modifying the determinants of α -ketoacid substrate selectivity in mycobacterium tuberculosis α -isopropylmalate synthase. *FEBS letters*, 588(9), 1603-1607
- Imber, R., Low, R. L., & Ray, D. S. (1983). Identification of a primosome assembly site in the region of the ori 2 replication origin of the *Escherichia coli* mini-F plasmid. *Proceedings of the National Academy of Sciences*, 80(23), 7132-7136.
- Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., & Boyer, H. W. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*, 198(4321), 1056-1063.

- Jensen, K. F. (1993). The *Escherichia coli* K12" wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *Journal of bacteriology*, 175(11), 3401-3407.
- Johnson, I. S. (1983). Human insulin from recombinant DNA technology. *Science*, 219(4585), 632-637.
- Jones, K. L., & Keasling, J. D. (1998). Construction and characterization of F plasmid-based expression vectors. *Biotechnology and bioengineering*, 59(6), 659-665.
- Jones, K. L., Kim, S. W., & Keasling, J. D. (2000). Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metabolic engineering*, 2(4), 328-338.
- Jürgen, B., Breitenstein, A., Urlacher, V., Büttner, K., Lin, H., Hecker, M., ... & Neubauer, P. (2010). Quality control of inclusion bodies in *Escherichia coli*. *Microbial cell factories*, 9(1), 41.
- Keasling, J. D. (1999). Gene-expression tools for the metabolic engineering of bacteria. *Trends in biotechnology*, 17(11), 452-460.
- Kelly, C. L., Liu, Z., Yoshihara, A., Jenkinson, S. F., Wormald, M. R., Otero, J. M., ... & Estevez, R. J. (2016). Synthetic chemical inducers and genetic decoupling enable orthogonal control of the rhaBAD promoter. *ACS synthetic biology*.
- Kessler, D., and J. Knappe. 1996. "Anaerobic Dissimilation of Pyruvate." *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. 1: 199–216. -> I CANNOT GET IT!
- Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L., & Keasling, J. D. (2001). Homogeneous expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology*, 147(12), 3241-3247.
- 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. *Journal of Bacteriology*, 182(24), 7029-7034.
- Khlebnikov, A., Skaug, T., & Keasling, J. D. (2002). Modulation of gene expression from the arabinose-inducible araBAD promoter. *Journal of industrial microbiology & biotechnology*, 29(1), 34-37.
- Kiefhaber, T., Rudolph, R., Kohler, H. H., & Buchner, J. (1991). Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Nature biotechnology*, 9(9), 825.
- Kiick, K. L., Weberskirch, R., & Tirrell, D. A. (2001). Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*. *Febs Letters*, 502(1-2), 25-30.
- Kisumi, M., Sugiura, M., & Chibata, I. (1976). Biosynthesis of norvaline, norleucine, and homoisoleucine in *Serratia marcescens*. *Journal of biochemistry*, 80(2), 333-339. (I)
- Kisumi, M., SUGIURA, M., Jyoji, K. A. T. O., & CHIBATA, I. (1976). L-Norvaline and L-homoisoleucine formation by *Serratia marcescens*. *Journal of biochemistry*, 79(5), 1021-1028. (II)
- Kohlhaw, G., Leary, T. R., & Umbarger, H. E. (1969). α -Isopropylmalate Synthase from *Salmonella typhimurium* PURIFICATION AND PROPERTIES. *Journal of Biological Chemistry*, 244(8), 2218-2225

- Komori, H., Matsunaga, F., Higuchi, Y., Ishiai, M., Wada, C., & Miki, K. (1999). Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution. *The EMBO journal*, 18(17), 4597-4607.
- Kreth, J., Lengeler, J. W., & Jahreis, K. (2013). Characterization of pyruvate uptake in *Escherichia coli* K12. *PLoS One*, 8(6), e67125.
- Laird, M. W., & Veeravalli, K. (2013). Methods and compositions for preventing norleucine mis-incorporation into proteins. US Patent Application WO 2014/047311 A1.
- Lawther RP, Calhoun DH, Adams CW, Hauser CA, Gray J, Hatfield GW. 1981. Molecular basis of valine resistance in *Escherichia coli* K12. *Proc. NatL Acad. Sci. USA* 78(2): 922-925.
- Lee, S. K., Chou, H. H., Pfeleger, B. F., Newman, J. D., Yoshikuni, Y., & Keasling, J. D. (2007). Directed evolution of AraC for improved compatibility of arabinose-and lactose-inducible promoters. *Applied and environmental microbiology*, 73(18), 5711-5715.
- Lee, S. Y. (1996). High cell-density culture of *Escherichia coli*. *Trends in biotechnology*, 14(3), 98-105.
- Lengeler, J. W., Drews, G., & Schlegel, H. G. (Eds.). (1999). *Biology of the Prokaryotes*. Georg Thieme Verlag.
- Li, H., Wang, B. S., Li, Y. R., Zhang, L., Ding, Z. Y., Gu, Z. H., & Shi, G. Y. (2017). Metabolic engineering of *Escherichia coli* W3110 for the production of L-methionine. *Journal of Industrial Microbiology & Biotechnology*, 1-14.
- Limberg MH, Pooth V, Wiechert W, 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*. *Eng. Life Sci.* 16, 610–619
- Lin, H. Y., Hanschke, R., Nicklisch, S., Riemschneider, S., Meyer, S., Gupta, A., ... & Schwahn, C. (2001). Cellular responses to strong overexpression of recombinant genes in *Escherichia coli*. In *Recombinant Protein Production with prokaryotic and eukaryotic cells. A comparative view on host physiology* (pp. 55-73). Springer, Dordrecht.
- Lin, R., d'Ari, R., & Newman, E. B. (1992). Lambda placMu insertions in genes of the leucine regulon: extension of the regulon to genes not regulated by leucine. *Journal of bacteriology*, 174(6), 1948-1955.
- Link, A. J., Mock, M. L., & Tirrell, D. A. (2003). Non-canonical amino acids in protein engineering. *Current opinion in biotechnology*, 14(6), 603-609.
- Liu, H., Kang, J., Qi, Q., & Chen, G. (2011). Production of lactate in *Escherichia coli* by redox regulation genetically and physiologically. *Applied biochemistry and biotechnology*, 164(2), 162-169.
- Loveland, A. B., Bah, E., Madireddy, R., Zhang, Y., Brilot, A. F., Grigorieff, N., & Korostelev, A. A. (2016). Ribosome• RelA structures reveal the mechanism of stringent response activation. *Elife*, 5, e17029.
- Lu, H. S., Tsai, L. B., Kenney, W. C., & Lai, H. (1988). Identification of unusual replacement of methionine by norleucine in recombinant interleukin-2 produced by *E. coli*. *Biochemical and biophysical research communications*, 156(2), 807-813.

- Madigan, M., J. Martinko, K. Bender, D. Buckley, and D. Stahl. 2014. Brock Biology of Microorganisms. Pearson Education Limited. doi:10.2460/ajvr.75.7.613.
- Madyagol, M., Al-Alami, H., Levarski, Z., Drahovská, H., Turňa, J., & Stuchlík, S. (2011). Gene replacement techniques for *Escherichia coli* genome modification. *Folia microbiologica*, 56(3), 253-263.
- Martínez-Alonso, M., González-Montalbán, N., García-Fruitós, E., & Villaverde, A. (2009). Learning about protein solubility from bacterial inclusion bodies. *Microbial cell factories*, 8(1), 4.
- Martinis SA, Fox GE. Non-standard amino acid recognition by *Escherichia coli* leucyl-tRNA synthetase. *Nucleic Acids Symp Ser*. 1997;36:125–128
- Mayer, M., & Buchner, J. (2004). Refolding of inclusion body proteins. In *Molecular diagnosis of infectious diseases* (pp. 239-254). Humana Press.
- Megerle, J. A., Fritz, G., Gerland, U., Jung, K., & Rädler, J. O. (2008). Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophysical journal*, 95(4), 2103-2115.
- Miller, H. I., Kikuchi, A., Nash, H. A., Weisberg, R. A., & Friedman, D. I. (1979). Site-specific recombination of bacteriophage λ : the role of host gene products. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 43, pp. 1121-1126). Cold Spring Harbor Laboratory Press.
- Miyazawa, T., Yokoyama, S., & Miyake, T. (1989). Method for producing protein containing nonprotein amino acids. U.S. Patent No. 4,879,223.
- Mnatsakanyan, N., K. Bagramyan, and A. Trchounian. 2004. "Hydrogenase 3 but Not Hydrogenase 4 Is Major in Hydrogen Gas Production by *Escherichia Coli* Formate Hydrogenlyase at Acidic pH and in the Presence of External Formate." *Cell Biochemistry and Biophysics* 41 (3): 357–66. doi:10.1385/CBB:41:3:357.
- Mogk, A., Kummer, E., & Bukau, B. (2015). Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation. *Frontiers in molecular biosciences*, 2, 22.
- Monod, J., Wyman, J., & Changeux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* (1965) 12, 88-118
- Morgan-Kiss, R. M., Wadler, C., & Cronan, J. E. (2002). Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. *Proceedings of the National Academy of Sciences*, 99(11), 7373-7377.
- Mori, H., Kondo, A., Ohshima, A., Ogura, T., & Hiraga, S. (1986). Structure and function of the F plasmid genes essential for partitioning. *Journal of molecular biology*, 192(1), 1-15.
- Munier, R., & Cohen, G. N. (1956). Incorporation d'analogues structuraux d'aminoacides dans les protéines bactériennes. *Biochimica et Biophysica Acta*, 21(3), 592-593.
- Muramatsu, R., Misawa, S., & Hayashi, H. (2003). Finding of an isoleucine derivative of a recombinant protein for pharmaceutical use. *Journal of pharmaceutical and biomedical analysis*, 31(5), 979-987.
- Muramatsu, R., Negishi, T., Mimoto, T., Miura, A., Misawa, S., & Hayashi, H. (2002). Existence of β -methylnorleucine in recombinant hirudin produced by *Escherichia coli*. *Journal of biotechnology*, 93(2), 131-142.

- Murphy, K. C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *Journal of bacteriology*, 180(8), 2063-2071.
- Naider, F., Z. Bohak, and J. Yariv. (1972). Reversible alkylation of a methionine residue near the active site of β -galactosidase. *Biochemistry* 11:3202-3208.
- Nakamori, S., Kobayashi, S., Nishimura, T., & Takagi, H. (1999). Mechanism of L-methionine overproduction by *Escherichia coli*: the replacement of Ser-54 by Asn in the MetJ protein causes the derepression of L-methionine biosynthetic enzymes. *Applied microbiology and biotechnology*, 52(2), 179-185.
- Nakashima, N., & Miyazaki, K. (2014). Bacterial cellular engineering by genome editing and gene silencing. *International journal of molecular sciences*, 15(2), 2773-2793.
- Nandi, P. & Sen, G.P. (1953). An antifungal substance from a strain of *B. subtilis*. *Nature* 171, 871-872
- Neidhardt, F. C., & Umbarger, H. E. (1987). Chemical composition of *E. coli*. *E. coli and S. typhimurium: Cellular and Molecular Biology*.
- Ni, J., Gao, M., James, A., Yao, J., Yuan, T., Carpick, B., ... & Farrell, P. (2015). Investigation into the mis-incorporation of norleucine into a recombinant protein vaccine candidate. *Journal of industrial microbiology & biotechnology*, 42(6), 971-975.
- Nielsen, J., & Keasling, J. D. (2016). Engineering Cellular Metabolism. *Cell*, 164(6), 1185-1197.
- Ning, Y., Wu, X., Zhang, C., Xu, Q., Chen, N., & Xie, X. (2016). Pathway construction and metabolic engineering for fermentative production of ectoine in *Escherichia coli*. *Metabolic engineering*, 36, 10-18.
- Novick, A., & Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7), 553.
- Olins, P. O., & Rangwala, S. H. (1989). A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in *Escherichia coli*. *Journal of Biological Chemistry*, 264(29), 16973-16976.
- Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., & Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, 427(6976), 737-740.
- Packeiser, H., Lim, C., Balagurunathan, B., Wu, J., & Zhao, H. (2013). An extremely simple and effective colony PCR procedure for bacteria, yeasts, and microalgae. *Applied biochemistry and biotechnology*, 169(2), 695-700.
- Parekh, B. S., & Hatfield, G. W. (1997). Growth rate-related regulation of the ilvGMEDA operon of *Escherichia coli* K12 is a consequence of the polar frameshift mutation in the ilvG gene of this strain. *Journal of bacteriology*, 179(6), 2086-2088.
- Parekh, B. S., Sheridan, S. D., & Hatfield, G. W. (1996). Effects of integration host factor and DNA supercoiling on transcription from the ilvPG promoter of *Escherichia coli*. *Journal of Biological Chemistry*, 271(34), 20258-20264.
- Patte JC (1996) Biosynthesis of threonine and lysine. In: Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (eds) *Escherichia coli and Salmonella: cellular and molecular biology*, 2nd edn. ASM, Washington, pp 528–541

- Pei, J., Dong, P., Wu, T., Zhao, L., Fang, X., Cao, F., ... & Yue, Y. (2016). Metabolic Engineering of *Escherichia coli* for Astragalin Biosynthesis. *Journal of Agricultural and Food Chemistry*, 64(42), 7966-7972.
- Phue, J. N., Noronha, S. B., Hattacharyya, R., Wolfe, A. J., & Shiloach, J. (2005). Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. *Biotechnology and Bioengineering*, 90(7), 805-820.
- Pickering, M., & Newton, P. (1990). Amino acid hydrolysis: Old problems, new solutions. *Lc Gc*, 8(10), 778-781.
- Platko, J. V., D. A. Willins, and J. M. Calvo. 1990. The *ilvIH* operon of *Escherichia coli* is positively regulated. *J. Bacteriol.* 172:4563–4570.
- Poulsen, P., Jensen, K. F., VALENTIN-HANSEN, P., Carlsson, P., & Lundberg, L. G. (1983). Nucleotide sequence of the *Escherichia coli* *pyrE* gene and of the DNA in front of the protein-coding region. *The FEBS Journal*, 135(2), 223-229.
- Pribnow, D. (1975). Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. *Journal of molecular biology*, 99(3), 419-443.
- Quail, M. A., Haydon, D. J., & Guest, J. R. (1994). The *pdhR*–*aceEF*–*lpd* operon of *Escherichia coli* expresses the pyruvate dehydrogenase complex. *Molecular microbiology*, 12(1), 95-104.
- Ramón, A., Señorale, M., & Marín, M. (2014). Inclusion bodies: not that bad.... *Frontiers in microbiology*, 5, 56.
- Randhawa, Z. I., Witkowska, H. E., Cone, J., Wilkins, J. A., Hughes, P., Yamanishi, K., ... & Arthur, P. (1994). Incorporation of norleucine at methionine positions in recombinant human macrophage colony stimulating factor (M-CSF, 4-153) expressed in *Escherichia coli*: structural analysis. *Biochemistry*, 33(14), 4352-4362.
- Reitz C. 2011: In-vivo screening system for engineered Triosephosphate Isomerase libraries. TU Berlin, Diploma Thesis.
- Reitz, C. (2017). Impacts of oscillating cultivation conditions on the quality of recombinant inclusion bodies in *Escherichia coli*. PhD Thesis.
- Reitz, C., Fan, Q., & Neubauer, P. (2018). Synthesis of non-canonical branched-chain amino acids in *Escherichia coli* and approaches to avoid their incorporation into recombinant proteins. *Current opinion in biotechnology*, 53, 248-253.
- Rhee KY, Opel M, Ito E, Hung S, Arfin SM, Hatfield GW. 1999. Transcriptional coupling between the divergent promoters of a prototypic *LysR*-type regulatory system, the *ilvYC* operon of *Escherichia coli*. *Proc Natl Acad Sci USA* 96:14294–14299.
- Rhee, K. Y., Parekh, B. S., & Hatfield, G. W. (1996). Leucine-responsive regulatory protein-DNA interactions in the leader region of the *ilvGMEDA* operon of *Escherichia coli*. *Journal of Biological Chemistry*, 271(43), 26499-26507.
- Rhee, K.Y., Senear, D.F. & Hatfield, G.W. (1998) Activation of gene expression by a ligand-induced conformational change of a protein–DNA complex. *J Biol Chem* 273: 11257–11266.

- Riesenber, D. (1991). High-cell-density cultivation of *Escherichia coli*. *Current Opinion in Biotechnology*, 2(3), 380-384.
- Riesenber, D., Menzel, K., Schulz, V., Schumann, K., Veith, G., Zuber, G., & Knorre, W. A. (1990). High cell density fermentation of recombinant *Escherichia coli* expressing human interferon alpha 1. *Applied microbiology and biotechnology*, 34(1), 77-82.
- Rinas, U., & Bailey, J. E. (1992). Protein compositional analysis of inclusion bodies produced in recombinant *Escherichia coli*. *Applied microbiology and biotechnology*, 37(5), 609-614.
- Rinas, U., Garcia-Fruitós, E., Corchero, J. L., Vázquez, E., Seras-Franzoso, J., & Villaverde, A. (2017). Bacterial inclusion bodies: discovering their better half. *Trends in biochemical sciences*, 42(9), 726-737.
- Roderick, S. L. (2005). The lac operon galactoside acetyltransferase. *Comptes rendus biologies*, 328(6), 568-575.
- Roesser, J. R., Nakamura, Y., & Yanofsky, C. (1989). Regulation of basal level expression of the tryptophan operon of *Escherichia coli*. *Journal of Biological Chemistry*, 264(21), 12284-12288.
- Salmon, K. A., Yang, C. R., & Hatfield, G. W. (2006). Biosynthesis and Regulation of the Branched-Chain Amino Acids†. *EcoSal Plus*, 2(1).
- Sawers, R. G., Blokesch, M., & Böck, A. (2004). Anaerobic Formate and Hydrogen Metabolism. *EcoSal Plus*, 1(1).
- Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends in Genetics*, 16(12), 559-565.
- Schweder, T., Krüger, E., Xu, B., Jürgen, B., Blomsten, G., Enfors, S. O., & Hecker, M. (1999). Monitoring of genes that respond to process-related stress in large-scale bioprocesses. *Biotechnology and bioengineering*, 65(2), 151-159.
- Schweder, T., Lin, H., Jürgen, B., Breitenstein, A., Riemschneider, S., Khalameyzer, V., ... & Neubauer, P. (2002). Role of the general stress response during strong overexpression of a heterologous gene in *Escherichia coli*. *Applied microbiology and biotechnology*, 58(3), 330-337.
- Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *Journal of bacteriology*, 189(23), 8746-8749.
- Shiloach, J., & Fass, R. (2005). Growing *E. coli* to high cell density—a historical perspective on method development. *Biotechnology advances*, 23(5), 345-357.
- Siegele, D. A., & Hu, J. C. (1997). Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences*, 94(15), 8168-8172.
- Silverstone, A. E., Arditti, R. R., & Magasanik, B. (1970). Catabolite-insensitive revertants of lac promoter mutants. *Proceedings of the National Academy of Sciences*, 66(3), 773-779.
- Sirko, A., Zehelein, E., Freundlich, M., & Sawers, G. (1993). Integration host factor is required for anaerobic pyruvate induction of pfl operon expression in *Escherichia coli*. *Journal of bacteriology*, 175(18), 5769-5777.

- Sletta, H., Nedal, A., Aune, T. E. V., Hellebust, H., Hakvåg, S., Aune, R., ... & Brautaset, T. (2004). Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in *Escherichia coli*. *Applied and environmental microbiology*, 70(12), 7033-7039.
- Soini, J., Falschlehner, C., Liedert, C., Bernhardt, J., Vuoristo, J., & Neubauer, P. (2008). Norvaline is accumulated after a down-shift of oxygen in *Escherichia coli* W3110. *Microb. Cell Fact.*, 7: 1-14. (I)
- Soini, J., Ukkonen, K., & Neubauer, P. (2008). High cell density media for *Escherichia coli* are generally designed for aerobic cultivations—consequences for large-scale bioprocesses and shake flask cultures. *Microbial Cell Factories*, 7(1), 26. (II)
- Soini, J., Ukkonen, K., & Neubauer, P. (2011). Accumulation of amino acids deriving from pyruvate in *Escherichia coli* W3110 during fed-batch cultivation in a two-compartment scale-down bioreactor. *Advances in Bioscience and Biotechnology*, 2(05), 336.
- Soper, T. S., Doellgast, G. J., & Kohlhaw, G. B. (1976). Mechanism of feedback inhibition by leucine: Purification and properties of a feedback-resistant α -isopropylmalate synthase. *Archives of biochemistry and biophysics*, 173(1), 362-374.
- Sousa, C., de Lorenzo, V., & Cebolla, A. (1997). Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology*, 143(6), 2071-2078.
- Steinsiek, S., & Bettenbrock, K. (2012). Glucose transport in *Escherichia coli* mutant strains with defects in sugar transport systems. *Journal of bacteriology*, 194(21), 5897-5908.
- Sugiura, M., Kisumi, M., & Chibata, I. (1981). Biosynthetic pathway of β -methylnorleucine, an antimetabolite produced by *Serratia marcescens*. *The Journal of antibiotics*, 34(10), 1283-1289.
- Sunasara, K. M., Cramer, S. M., Hauer, C. R., Rupp, R. G., & Shoup, V. A. (1999). Characterization of recombinant human brain-derived neurotrophic factor variants. *Archives of biochemistry and biophysics*, 372(2), 248-260.
- Sutton, A., & Freundlich, M. (1980). Regulation by cyclic AMP of the *ilvB*-encoded biosynthetic acetohydroxy acid synthase in *Escherichia coli* K12. *Molecular and General Genetics MGG*, 178(1), 179-183.
- Sycheva EV, Yampol'skaya TA, Preobrajenskaya ES, Novikova AE, Matrosov NG, Stoyanova NV. 2007. Overproduction of Noncanonical Amino Acids by *Escherichia coli* Cells. *Microbiology* 76(6): 712-718.
- Széliová, D., Krahulec, J., Šafránek, M., Lišková, V., & Turňa, J. (2016). Modulation of heterologous expression from P BAD promoter in *Escherichia coli* production strains. *Journal of Biotechnology*, 236, 1-9.
- Tang, Y., & Tirrell, D. A. (2002). Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry*, 41(34), 10635-10645.
- Tchetina, E., & Newman, E. B. (1995). Identification of Lrp-regulated genes by inverse PCR and sequencing: regulation of two *mal* operons of *Escherichia coli* by leucine-responsive regulatory protein. *Journal of bacteriology*, 177(10), 2679-2683.

- Tedin, K., & Norel, F. (2001). Comparison of Δ relA strains of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium suggests a role for ppGpp in attenuation regulation of branched-chain amino acid biosynthesis. *Journal of bacteriology*, 183(21), 6184-6196.
- Tegel, H., Ottosson, J., & Hober, S. (2011). Enhancing the protein production levels in *Escherichia coli* with a strong promoter. *FEBS journal*, 278(5), 729-739.
- Teich, A., Meyer, S., Lin, H. Y., Andersson, L., Enfors, S. O., & Neubauer, P. (1999). Growth rate related concentration changes of the starvation response regulators σ S and ppGpp in glucose-limited fed-batch and continuous cultures of *Escherichia coli*. *Biotechnology progress*, 15(1), 123-129.
- Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*, 72(2), 211-222.
- Thomason, L. C., Sawitzke, J. A., Li, X., Costantino, N., & Court, D. L. (2014). Recombineering: genetic engineering in bacteria using homologous recombination. *Current protocols in molecular biology*, 106(1), 1-39.
- Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., Hightower, G. A., Smith, J. T., & Conway, T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular microbiology*, 68(5), 1128-1148.
- Tsai, et al. "Control of mis-incorporation of de novo synthesized norleucine into recombinant interleukin-2 in *E. coli*." *Biochemical and biophysical research communications* 156.2 (1988): 733-739.
- Tsui, P., & Freundlich, M. (1990). Integration host factor bends the DNA in the *Escherichia coli* ilvBN promoter region. *Molecular and General Genetics MGG*, 223(2), 349-352.
- Umbarger H E. Biosynthesis of the branched chain amino acids. In: Neidhardt F C, et al., editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology. 2nd ed. Vol. 1. Washington, D.C.: American Society for Microbiology; 1996. pp. 442–457.
- Umbarger, H. E. (1956). Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *American Association for the Advancement of Science*.
- Umbarger, H. E., & Brown, B. (1957). Threonine deamination in *Escherichia coli* ii.: Evidence for two L-threonine deaminases¹. *Journal of bacteriology*, 73(1), 105.
- Usuda, Y., & Kurahashi, O. (2005). Effects of deregulation of methionine biosynthesis on methionine excretion in *Escherichia coli*. *Applied and environmental microbiology*, 71(6), 3228-3234.
- Vallejo, L. F., & Rinas, U. (2004). Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial cell factories*, 3(1), 11.
- Van Deventer, J. A., Fisk, J. D., & Tirrell, D. A. (2011). Homoisoleucine: a translationally active leucine surrogate of expanded hydrophobic surface area. *ChemBioChem*, 12(5), 700-702.
- Vartak, N. B., Liu, L., Wang, B. M., & Berg, C. M. (1991). A functional leuABCD operon is required for leucine synthesis by the tyrosine-repressible transaminase in *Escherichia coli* K12. *Journal of bacteriology*, 173(12), 3864-3871.

- Veeravalli, K., Laird, M. W., Fedesco, M., Zhang, Y., & Yu, X. C. (2015). Strain engineering to prevent norleucine incorporation during recombinant protein production in *Escherichia coli*. *Biotechnology progress*, 31(1), 204-211.
- Via, P., Badia, J., Baldomá, L., Obradors, N., & Aguilar, J. (1996). Transcriptional regulation of the *Escherichia coli* rhaT gene. *Microbiology*, 142(7), 1833-1840.
- Vinogradov, V., Vyazmensky, M., Engel, S., Belenky, I., Kaplun, A., Kryukov, O., ... & Chipman, D. M. (2006). Acetohydroxyacid synthase isozyme I from *Escherichia coli* has unique catalytic and regulatory properties. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1760(3), 356-363.
- Vitreschak, A. G., Lyubetskaya, E. V., Shirshin, M. A., Gelfand, M. S., & Lyubetsky, V. A. (2006). Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS microbiology letters*, 234(2), 357-370.
- Wahren, J., Ekberg, K., Johansson, J., Henriksson, M., Pramanik, A., Johansson, B. L., ... & Jörnvall, H. (2000). Role of C-peptide in human physiology. *American Journal of Physiology-Endocrinology And Metabolism*, 278(5), E759-E768.
- Walsh, G. (2014). Biopharmaceutical benchmarks 2014. *Nature biotechnology*, 32(10), 992-1000.
- Wang, Q., Wu, J., Friedberg, D., Plakto, J., & Calvo, J. M. (1994). Regulation of the *Escherichia coli* lrp gene. *Journal of bacteriology*, 176(7), 1831-1839.
- Wegerer, A., Sun, T., & Altenbuchner, J. (2008). Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC biotechnology*, 8(1), 2.
- Whalen, W. A., & Berg, C. M. (1982). Analysis of an avtA:: Mu d1 (Ap lac) mutant: metabolic role of transaminase C. *Journal of bacteriology*, 150(2), 739-746.
- Wiegel, J., & Schlegel, H. G. (1977). α -Isopropylmalate synthase from *Alcaligenes eutrophus* H 16. *Archives of microbiology*, 112(3), 239-246.
- Williams, D. C., Van Frank, R. M., Muth, W. L., & Burnett, J. P. (1982). Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science*, 215(4533), 687-689.
- Winther-Larsen, H. C., Blatny, J. M., Valand, B., Brautaset, T., & Valla, S. (2000). Pm promoter expression mutants and their use in broad-host-range RK2 plasmid vectors. *Metabolic engineering*, 2(2), 92-103.
- Wolfe, A. J. (2005). The acetate switch. *Microbiology and molecular biology reviews*, 69(1), 12-50.
- Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., & Koffas, M. A. (2016). Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in biotechnology*
- Xu B, Jahic M, Blomsten G, Enfors SO. 1999a. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. *Applied Microbiology and Biotechnology* 51(5):564-571.
- Yoon SH, Han MJ, Jeong H, Lee CH, Xia XX, Lee DH, Shim JH, Lee SY, Oh TK, Kim JF. 2012. Comparative multi-omics systems analysis of *Escherichia coli* strains B and K12. *Genome. Biol.* 13: R37.

- Young, T. S., & Schultz, P. G. (2010). Beyond the canonical 20 amino acids: expanding the genetic lexicon. *Journal of Biological Chemistry*, 285(15), 11039-11044.
- Yu, X., Wang, X., & Engel, P. C. (2014). The specificity and kinetic mechanism of branched-chain amino acid aminotransferase from *Escherichia coli* studied with a new improved coupled assay procedure and the enzyme's potential for biocatalysis. *The FEBS journal*, 281(1), 391-400.
- Yuan, T., & Vogel, H. J. (1999). Substitution of the methionine residues of calmodulin with the unnatural amino acid analogs ethionine and norleucine: biochemical and spectroscopic studies. *Protein science*, 8(1), 113-121.
- Zhu, S., Gong, C., Ren, L., Li, X., Song, D., & Zheng, G. (2013). A simple and effective strategy for solving the problem of inclusion bodies in recombinant protein technology: His-tag deletions enhance soluble expression. *Applied microbiology and biotechnology*, 97(2), 837-845.
- Zündorf I, Dingermann T. 2001. Bereitstellung ausreichender Mengen von Humaninsulin. Vom Rinder-, Schweine-, Pferde-Insulin zum Humaninsulin: Die biotechnische und gentechnische Insulin-Herstellung. *Pharmazie in unserer Zeit* 30: 27-32.

Websites

Website 1:

<http://www.vectronbiosolutions.com/technologies/>

Website 2:

<https://nebiocalculator.neb.com/#!/ligation>

Website 3:

http://www.clontech.com/US/Support/xxclt_onlineToolsLoad.jsp?citeId=http://bioinfo.clontech.com/infusion/molarRatio.do§ion=16260&xxheight=750

Website 4:

https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/LSG/manuals/pbaddest49_man.pdf

Website 5:

http://2011.igem.org/Team:DTU-Denmark/Project_improving_araBAD

9. Theses

- Genetic regulation of target genes involved in the BCAA biosynthetic pathway affects biosynthesis and mis-incorporation of ncBCAA into recombinant proteins. Up-regulation of *ilvC*, *ilvIH* or *ilvGM* and down-regulation of *leuA* or *ilvBN* triggers a reduction of ncBCAA biosynthesis and mis-incorporation. Up-regulation of single operons *ilvIH* and *ilvGM* showed the highest reduction of ncBCAA biosynthesis and mis-incorporation.
- The pyruvate pathway is more likely to redirect more flux to α -ketobutyrate than the threonine pathway.
- Norvaline, norleucine and by-products of the overflow metabolism accumulate in *E. coli* under cultivation conditions subjected to pyruvate pulsing combined with oxygen limitation.
- Metabolic effects triggered in *E. coli* cultivations subjected to pyruvate pulsing and O₂ limitation occur first in the cytosol (intracellular soluble protein fraction), where ncBCAAs are synthesized. Afterwards, translation machinery mis-incorporates those previously synthesized ncBCAA present in the cytosol into the nascent recombinant proteins (inclusion body fraction).
- Norleucine concentration is significantly higher than norvaline in the inclusion body fraction while β -methylnorleucine could not be detected. This suggests that the amount of a certain cBCAA in the amino acid sequence of a recombinant protein is not the main factor determining the mis-incorporation probability of the respective ncBCAA analog in such recombinant protein.
- Despite being norvaline and norleucine present in similar concentration levels intracellularly, the probability of misaminoacylation by met-tRNA seems to be higher than by leu-tRNA.
- Norvaline and norleucine concentrations were higher at 15 L reactor than at mini-reactor scale. Although not comparable to industrial scale reactors, it might be that mixing in the 15L reactor is inefficient, thus triggering formation of O₂ and glucose gradients, which, in turn, lead to ncBCAA formation, even under standard cultivation conditions.
- IPTG-mediated induction of recombinant protein expression is the main factor triggering ncBCAA formation under standard cultivation conditions.
- Neither genetic regulation of target genes involved in the BCAA biosynthetic pathway nor cultivation conditions subjected to pyruvate pulsing and oxygen limitation caused a significant alteration of β -methylnorleucine levels in *E. coli*. This may suggest that β -methylnorleucine is synthesized by an alternative unknown metabolic pathway.
- For each target gene, different induction strength, i.e. L-arabinose concentration, is necessary in order to trigger genetic expression levels enough to recover cell growth levels of the wild type strain. This might be explained due to the different degree of essentiality of the target genes for the cell. Lack of one of the three *ilvGM*, *ilvIH* and *ilvBN* gene products does not dramatically affect growth behavior since, as isoenzymes, its function can be also carried out by the other two available gene products.
- Induction efficiency of lac-based promoters present in expression plasmids is strongly dependent on the levels of LacI repressor endogenously expressed in the *E. coli* strain selected as host for expression.

10. Appendix

Table S1. Reagents used in this study.

Category	Name	Catalogue number	Manufacturer
Antibiotics	Ampicillin sodium salt	A9518	Sigma
	Chloramphenicol	C0378	Sigma
	Kanamycin disulfate salt	K1876	Sigma
Amino acids	L-norleucine	N6877	Sigma
	L-norvaline	N7627	Sigma
	β-methylnorleucine	4036488	Bachem
	L-2-aminobutyric acid	A1879	Sigma
Enzymes	Amylase Reagent A 3000 U/L	-	Biosilta
	Lysonase Bioprocessing Reagent	71230	Merck
	Dreamtaq polymerase	EP0702	Thermoscientific
	Phusion DNA Polymerase	M0530	NEB
	Pfu polymerase	EP0571	Thermoscientific
	XhoI	R0146	NEB
	MssI	R0560	NEB
	NheI-HF	R3131	NEB
	NotI-HF	R3189	NEB
	PstI-HF	R3140	NEB
	HindIII-HF	R3104	NEB
	DpnI	R0176	NEB
	EcoRI-HF	R3101	NEB
	PaeI-HF	R3182	NEB
	PciI	R0655	NEB
	Quick ligase	M2200	NEB
	Exo-Star Reagent	GEUS78210	Sigma
Protein and DNA	Quick-Load® 2-Log DNA Ladder	N0469S	NEB

markers	SeeBlue® Plus2 Pre-stained Protein Standard	LC5925	Thermo Fisher Scientific
Protein and nucleic acid stains	Ethidium bromide, 10 mg/mL	15585-011	Invitrogen
	GelRed Nucleic Acid Gel Stain 10,000X stock solution	41003	Biotium
	Instant Blue	ISB1L	Sigma-Aldrich
SDS-PAGE reagents	NuPAGE LDS Sample Buffer (4X)	NP0007	Thermo Fisher Scientific
	NuPAGE MES SDS Running Buffer (20X)	NP0002	Thermo Fisher Scientific
	NuPAGE Sample Reducing Agent (10X)	NP0009	Thermo Fisher Scientific
	NuPAGE Antioxidant	NP0005	Thermo Fisher Scientific
	Bolt LDS Sample Buffer (4X)	B0008	Thermo Fisher Scientific
	Bolt MES SDS Running Buffer (20X)	B0002	Thermo Fisher Scientific
	Bolt Sample Reducing Agent (10X)	B0009	Thermo Fisher Scientific
	Bolt Antioxidant	B0005	Thermo Fisher Scientific
	NuPAGE 12% Bis-Tris Plus gels	NP0342BOX	Thermo Fisher Scientific
	Bolt 12% Bis-Tris Plus gels	NW00120BOX	Thermo Fisher Scientific
Components for DNA electrophoresis	6X Gel Loading Dye	B7024S	NEB
	E-Gel® 1.2 % Agarose (EtBr stained)	G501801	Thermoscientific
	Agarose	840004	Biozym
PCR reagents	Thermopol-buffer 10X	B9004	NEB
	10X DreamTaq buffer	B65	Thermoscientific
	5X Phusion GC Buffer	M0530	NEB
	5X Phusion HF Buffer	M0530	NEB
	100% DMSO	B0515A	NEB
	MgCl ₂	M0530	NEB

	dNTPs 10mM	N0447	NEB
	10 mM Tris/1 mM EDTA (TE buffer)	93283	Sigma
Buffers	10X CutSmart Buffer	B7204	NEB
	Quick ligase reaction buffer 2X	M2200	NEB
	10X TAE Buffer	15558-042	Gibco
	Quick ligation kit	M2200	NEB
Kits	In-Fusion® HD Cloning Kit	638909	Takara
	QIAGEN Plasmid Plus Midi Kit	ID12943	Qiagen
	QIAprep® Miniprep kit	27106	NEB
	QIAquick Gel Extraction Kit	28706	Qiagen
	QIAquick PCR Purification Kit	28106	Qiagen
	EZ:faast™ for free (physiological) amino acid analysis by GC-FID kit	KG0-7165	Phenomenex
	BugBuster Protein Extraction Reagent	70584-4	Merck
	D(+)-Glucose monohydrate	108342	Merck
Reagents for media preparation	MgSO4*7H2O	M2773	Sigma
	Invitrogen UltraPure™ Distilled Water	10977-035	Thermo Scientific
	LB-Agar (Lennox)	X965.1	Roth
	LB-Medium (Lennox)	X964.1	Roth
	SOC medium Outgrowth Medium	B9020S	NEB
	FeCl3*6H2O	236489	Sigma
	CaCl2*2H2O	C3306	Sigma
	MnCl2*4H2O	M8054	Sigma
	ZnSO4*7H2O	Z0251	Sigma
	CoCl2*6H2O	C8661	Sigma
	CuCl2*2H2O	459097	Sigma

	NiCl ₂ *6H ₂ O	13613	Sigma
	Na ₂ MoO ₄ *5H ₂ O	331058	Sigma
	H ₃ BO ₃	31146	Sigma
	K ₂ HPO ₄	1551128	Sigma
	KH ₂ PO ₄	NIST200B	Sigma
	Na citrate*2H ₂ O	W302600	Sigma
	(NH ₄) ₂ SO ₄	A4418	Sigma
	NaCl	S9625	Sigma
	Thiamine hydrochloride	T4625	Sigma
	Bacto™ Casamino acids	223050	BD
	Na ₂ HPO ₄ *2H ₂ O	1.06580.0500	Merck
	NH ₄ Cl	1.011143.0050	Merck
	Na ₂ SO ₄	31481	Sigma
	NaH ₂ PO ₄	S0751	Sigma
	(NH ₄) ₂ -H-citrate	09833	Sigma
	MnSO ₄ *H ₂ O	1.05941.0250	Merck
	CuSO ₄ *5H ₂ O	31293	Sigma
	L-arabinose	A3256	Sigma
	EnPump 200 substrate	-	Enpresso
	Sodium pyruvate	P2256	Sigma
	Desmophen	-	Covestro
	Antifoam 204	A6426	Sigma
	25 % NH ₄ OH solution	06010.4010	Bernd Kraft
	m-toluate	T36609	Sigma
	IPTG	I6758	Sigma
	Hydrochlorhydric acid 5M	10605882	Fischer Scientific
Others	Isopropanol	184130010	Acros Organics
	Glycerol	535036	Hedinger

Betaine	B24397	Alfa Aesar
Bardac	-	Lonza

Table S2. Materials used in this study.

Material	Catalog number	Manufacturer
Eppendorf tubes 1.5 mL	0030120.086	Eppendorf
Eppendorf tubes 2 mL	0030123.344	Eppendorf
Eppendorf tubes 5 mL	0030119.401	Eppendorf
PCR tubes 0.2 mL	0030124.332	Eppendorf
PCR tube Stripes 0.2 mL (used for sequencing)	0030124.359	Eppendorf
Nalgene Cryogenic vials (2 mL)	5000-0020	Thermo Scientific
Gene Pulser®/MicroPulser™ Electroporation Cuvettes, 0.2 cm gap	1652086	Biorad
15 mL round-bottom tube	0030122.151	Eppendorf
50 mL Falcon tubes	0030122.178	Eppendorf
24-well deep well plates (Pall mini-reactor)	MRT-PRC-21	Pall
96-well microplate	04-083-0150	Nerbe plus
Syringe filter 0.45 µm	10462100	Healthcare Life Sciences
Syringe filter 0.22 µm	10462200	GE Healthcare Life Sciences
Syringes	Omnifix Syringes	B.Braun
Canules	Sterican	B. Braun
Pipette tips	epT.I.P.S.	Eppendorf
Pipettes kit	Research plus	Eppendorf
Pipette filler	Accu-jet pro	Brand
Serological pipettes	Nunc serological pipette	Thermo Scientific
Filtration unit (diff. volumes)	Express Plus 0.22 µm (SCGPU11RE)	Millipore
Steriflip®	SCGP00525	Millipore
Cuvette PMMA, semi-micro (1.5 mL)	759115	Brand

Inoculation loop 1 µL	731165	Fischer Scientific
Inoculation loop 10 µL	731175	Fischer Scientific
Inoculation needle	731185	Fischer Scientific
L-shaped spreader	174CS05	Copan

Table S3. Equipment used in this study.

Equipment	Model	Manufacturer
Thermocycler	SimpliAmp Thermal Cycler	Thermo Scientific
Electroporator	Gene Pulser Xcell	Biorad
Spectrophotometer	NanoDrop One	Thermo Scientific
Photometer	Ultraspec 2100 pro	Amersham Bioscience
Gel documentation	Gel Doc EZ Imager	Biorad
Sonotrode	VialTweeter UP200St	Hielscher
GC-FID instrument	7890A	Agilent Technologies
GC Autoinjector	CombiPAL Injector CTC Analytics G6501-CTC	Agilent Technologies
Hydrogen generator	500	Schmidlin
GC column	ZB-AAA Zebron Amino Acid GC Column 10mx0.25mm CG0-7169	Phenomenex
Mini-reactor system	Pall Micro24	Microreactor Technologies Inc.
15L reactor	Type 880142.8, Nr. 209	Braun Melsungen
DCU reactor	BiostatR ED	B. Braun Biotech International
BioPAT Trace	biopattrace	Sartorius
Analytical balance	ME204T/00	Mettler Toledo
Balance	Laboratory LC6200S	Sartorius
Centrifuge	75005510/01	Biofuge Fresco
Centrifuge	Heraeus Multifuge X1R	Thermo Scientific
Centrifuge	Heraeus Pico 21	Thermo Scientific
Vortex	444-1372	VWR International
DNA electrophoresis units	E-Gel iBase™ Power System	Thermo Scientific

Thermomixer	88880028	Thermo Scientific
Vertical shaker	Trayster D	IKA
Incubator	IPP 55 ^{PLUS}	Memmert
Incubator	HERATHERM Incubator	Thermo Scientific
pH Meter	Multi 9310 IDS	WTW
Protein electrophoresis unit	XCell SureLoc Mini-Cell	Thermo Scientific
Horizontal gel electrophoresis chamber	460.000 (midi large)	-
Horizontal gel electrophoresis chamber	MSMINIDUO	Biozym
Water bath	WNB22	Memmert
UV Transilluminator	UST-20M-8R (Transilluminator BioView)	Biostep
Laminar flow bench	50067140	Heraeus
Magnetic stirrer	RSM-10HS	Phoenix Instrument
Microwaves	MM817ALR	Siemens
Incubator shaker	Kuhner SHAKER X / Climo-shaker ISF1-X	Kuhner
Incubator shaker	886342/3 (Certomat R)	B.

Table S4. Software used in this study.

Software	Provider	Application
SnapGene	GSL Biotech LLC	Primer design, generation of plasmid maps, alignment of sequences, <i>in silico</i> molecular verification.
NEBio® Claculator	NEB	Calculation of ligation parameters.
Tm Calculator	NEB	Determination of Ta of primers for PCR.
In-Fusion® Molar Ratio Calculator	Takara	Calculation optimal amounts of vector and insert for the In-Fusion Cloning reaction.
Multiple Primer Analyzer	Thermo Scientifc	Analysis and comparison of

		multiple primer sequences.
Lucullus PIMS v 3.5.2	Securecell	Process monitoring during fermentations.
Image Lab	Biorad	Acquisition and anlysis of DNA and protein gel images.
Agilent Chemstation, Rev. B.03.02 [341]	Agilent Technologies	Programming of GC runs and chromatogram analysis.
Leica Application Suite V4	Leica Microsystems	Acquisition, storage, annotation and display of microscope images.

Table S5. Strains used in this study and its features.

Strain	Genotype	Source	Comment
<i>E. coli</i> K-12 NEB5α	<i>fhuA2, (argF-lacZ)U169, phoA, glnV44, 80(lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17</i>	NEB (Cat. Nr.: C2987H)	Used as recipient strain for plasmid storage.
<i>E. coli</i> K-12 W3110M	F ⁻ , λ ⁻ , IN(<i>rrnD-rrnE</i>)1, <i>rph-1, ilvG-1, lacI^q</i>	Internal	<i>E. coli</i> wild type strain. Used as recipient strain.
<i>E. coli</i> K-12 BW25113	F ⁻ , Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(::rrnB-3), λ ⁻ , <i>rph-1, ilvG-1, Δ(rhaD-rhaB)</i> 568, <i>hsdR514, lacI⁺</i>	CGSC#7636	<i>E. coli</i> wild type strain mainly used in this study. Used as recipient strain.
<i>E. coli</i> K-12 BW25113 pKD46	See <i>E. coli</i> K-12 BW25113	CGSC#7739	Strain containing plasmid pKD46, necessary for the recombineering process used in this study.
<i>E. coli</i> K-12 BW25141 pKD3	F ⁻ , Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(::rrnB-3), Δ(<i>phoB-phoR</i>)580, λ ⁻ , <i>galU95, ΔuidA3::pir⁺, recA1, endA9(del-ins)::FRT, rph-1, Δ(rhaD-rhaB)</i> 568, <i>hsdR51</i>	CGSC#7631	Strain containing plasmid pKD3, necessary for the recombineering process used in this study.
<i>E. coli</i> K-12 BW25141 pKD4	See <i>E. coli</i> K-12 BW25141 pKD3	CGSC#7632	Strain containing plasmid pKD4, necessary for the recombineering process used in this study.

<i>E. coli</i> K-12 BT340 pCP20	F-, $\Delta(\text{argF-lac})169$, $\phi 80\text{dlacZ58(M15)}$, <i>glnX44(AS)</i> , λ -, <i>rfbC1</i> , <i>gyrA96(NalR)</i> , <i>recA1</i> , <i>endA1</i> , <i>spoT1</i> , <i>thiE1</i> , <i>hsdR17</i>	CGSC#7629	Strain containing plasmid pCP20, necessary for the recombineering process used in this study.
<i>E. coli</i> K-12 BW25113 <i>leuA:kanR</i> pKD46*	See <i>E. coli</i> K-12 BW25113	CGSC#JW0073-1	<i>leuA</i> KO-mutant containing plasmid pKD46. Kanamycin resistance cassette is integrated in <i>leuA</i> genetic region.
<i>E. coli</i> K-12 BW25113 <i>ilvA:kanR</i> pKD46*	See <i>E. coli</i> K-12 BW25113	CGSC#JW3745-2	<i>ilvA</i> KO-mutant containing plasmid pKD46. Kanamycin resistance cassette is integrated in <i>ilvA</i> genetic region.
<i>E. coli</i> K-12 BW25113 <i>ilvC:kanR</i> pKD46*	See <i>E. coli</i> K-12 BW25113	CGSC#JW3747-2	<i>ilvC</i> KO-mutant containing plasmid pKD46. Kanamycin resistance cassette is integrated in <i>ilvC</i> genetic region.
<i>E. coli</i> K-12 BW25113 <i>thrA:kanR</i> pKD46*	See <i>E. coli</i> K-12 BW25113	CGSC#JW0001-1	<i>thrA</i> KO-mutant containing plasmid pKD46. Kanamycin resistance cassette is integrated in <i>thrA</i> genetic region.
<i>E. coli</i> K-12 BW25113 ΔilvIH	See <i>E. coli</i> K-12 BW25113	This study	<i>ilvIH</i> KO-mutant. Used as recipient strain.
<i>E. coli</i> K-12 BW25113 ΔilvBN	See <i>E. coli</i> K-12 BW25113	Sarah Charaf	<i>ilvBN</i> KO-mutant. Used as recipient strain.

*: These strains were further processed as follows in order to obtain the recipient strain: plasmid pKD46 was curated from the acquired *E. coli* K-12 BW25113 single knock-out mutants and the respective electro-competent cells were transformed with plasmid pCP20 in order to trigger removal of the kanamycin cassette from the genome by FRT-specific recombination. Plasmid pCP20 was then curated giving the corresponding *leuA*, *thrA*, *ilvA* and *ilvC* KO-recipient strains.

Table S6. Plasmids used in this study and its features.

Plasmid	Size (bp)	Antibiotic resistance	Source	Expressed protein	Features
pKD3	2804	amp, cm	CGSC#7631	-	R6K ori, 3 priming sites for PCR, cmR flanked by FRT sites

pKD4	3267	amp, kan	CGSC#7632	-	R6K ori, 3 priming sites for PCR, kanR flanked by FRT sites
pKD46	6329	amp	CGSC#7739	lambda (λ) Red recombination system (Proteins Exo, Bet und Gam)	<i>araBAD</i> promoter, pSC101 ori (Rep101)
pCP20	9497	amp, cm	CGSC#7629	Flippase (for FRT-specific recombination)	pSC101 ori (Rep101)
pSW3	3609	amp	Internal	Mini-proinsulin	P _{tac} promoter, pMB1 (derivative) ori
pSW3_ <i>lacI</i> ⁺	3982	amp	This study	Mini-proinsulin and LacI repressor	P _{tac} promoter, pMB1 (derivative) ori, LacI (<i>lacI</i> ⁺ promoter)
pSW3_ <i>lacI</i> ^q	3982	amp	This study	Mini-proinsulin and LacI repressor	P _{tac} promoter, pMB1 (derivative) ori, LacI (<i>lacI</i> ^q promoter)
pET-coco1	12474	cm	Novagen (Cat. Nr.: 71129)	-	T7 promoter, oriV (TrfA, <i>araBAD</i> promoter), oriS (RepE, SopA, SopB, SopC)
pACG_ <i>araBAD</i>	7237	cm	This study	-	<i>araBAD</i> promoter, oriS (RepE, SopA, SopB, SopC)
pACG_ <i>araBAD</i> _ <i>thrA</i>	9680	cm	This study	Bifunctional aspartokinase/homoserine dehydrogenase 1	See pACG_ <i>araBAD</i>
pACG_ <i>araBAD</i> _ <i>lvA</i>	8762	cm	This study	L-threonine dehydratase biosynthetic	See pACG_ <i>araBAD</i>
pACG_ <i>araBAD</i> _ <i>lvA</i>	8789	cm	This study	2-isopropylmalate synthase	See pACG_ <i>araBAD</i>
pACG_ <i>araBAD</i> _ <i>lvIH</i>	9436	cm	This study	Acetolactate synthase isozyme 3	See pACG_ <i>araBAD</i>
pACG_ <i>araBAD</i> _ <i>lvBN</i>	9200	cm	This study	Acetolactate synthase isozyme 1	See pACG_ <i>araBAD</i>
pACG_ <i>araBAD</i> _ <i>lvGM</i>	9121	cm	This study	Acetolactate synthase isozyme 2	See pACG_ <i>araBAD</i>
pACG_ <i>araBAD</i> _ <i>lvC</i>	8693	cm	This study	Ketol-acid reductoisomerase (NADP(+))	See pACG_ <i>araBAD</i>

pACG_XylSPm	7948	cm	This study	-	XylSPm promoter, oriS (RepE, SopA, SopB, SopC)
pACG_XylSPm_thrA	10391	cm	This study	Bifunctional aspartokinase/homoserine dehydrogenase 1	See pACG_XylSPm
pACG_XylSPm_ilvA	9473	cm	This study	L-threonine dehydratase biosynthetic	See pACG_XylSPm
pACG_XylSPm_leuA	9500	cm	This study	2-isopropylmalate synthase	See pACG_XylSPm
pACG_XylSPm_ilvC	9404	cm	This study	Ketol-acid reductoisomerase (NADP(+))	See pACG_XylSPm
pACG_XylSPm_ilvBN	9911	cm	This study	Acetolactate synthase isozyme 1	See pACG_XylSPm
16ABZ5NP_193 4177_araBAD_MCS1	3876	kan	GeneArt	-	ColE1 ori
16ADFSUP_203 4902_lacI ^q _promotor	4214	kan	GeneArt	-	ColE1 ori
17ABNF2P_211 5088_sacBopt	3709	kan	GeneArt	Levansucrase (codon-optimized for <i>E. coli</i>)	ColE1 ori
16ADCJKP_202 8165_XylSPm	4461	kan	GeneArt	-	ColE1 ori

Table S7. Primers used in this study and its application.

Primer name	Sequence 5' → 3'	Primer application
pSW3_F1_seq	AACCTTTCGCGGTATGGCATG	1. Sequencing of plasmid pSW3. 2. Primers pSW3_R3_seq and pSW3_R6_seq were additionally used for plasmid verification by colony PCR.
pSW3_R1_seq	GGGCGCTATCATGCCATAC	
pSW3_R2_seq	TTGATGGGTGTCTGGTCAGAGAC	
pSW3_F2_seq	ATGAAGACGGTACGCGACTG	
pSW3_F3_seq	GCGTGGTATCGTTGAACAATG	
pSW3_R3_seq	CAACATTGTTCAACGATACCAC	
pSW3_F4_seq	GACTCAACTGCAACTGGAACAC	
pSW3_R4_seq	CGAAACGGCCCTCAATCATAC	

pSW3_F5_seq	CCGCTTTATCAGAAGCCAGAC	
pSW3_R5_seq	GCTACGTGACTGGGTCAT	
pSW3_F6_seq	CCAGCAAAAGGCCAGGAAC	
pSW3_R6_seq	GACCTACACCGAACTGAGATAC	
pSW3_F7_seq	TCTGCTGAAGCCAGTTACCTTC	
pSW3_R7_seq	GAGTCAGGCAACTATGGATGAAC	
pSW3_F8_seq	GCAGGCATCGTGGTGTAC	
Mut_ <i>lacI</i> ⁺ _F	P-CATCGAATGGCGCAAAACCTTTCGC	Mutagenesis of plasmid pSW3_ <i>lacI</i> ⁺ to generate pSW3_ <i>lacI</i> ⁺ .
Mut_ <i>lacI</i> ⁺ _R	P-GTGTGCGATGCCGCTTCG	
araBAD_InFusion_F	CTCCCGGCATCCGCTTACAGACAAG	PCR from plasmid 16ABZ5NP_1934177 to generate Fragment 1, which participates in the InFusion cloning process to generate plasmid pACG_araBAD.
araBAD_InFusion_R	CAAAAAACCCCTCAAGACCCGTTTAGAG	
cmR_InFusion_F	TTGAGGGGTTTTTTGATTAAATAGCGCTAACCGT TTTTATCAGGCTCTG	PCR from plasmid pCP20 to generate Fragment 2, which participates in the InFusion cloning process to generate plasmid pACG_araBAD.
cmR_InFusion_R	CACCTAGGCTCGAGTGATCGGCACGTAAGAGGTT CCAAC	
ori2_InFusion_F	ACTCGAGCCTAGGTGAGCGAGGAAGCACCAGGG AACAG	PCR from plasmid pETcoco1 to generate Fragment 3, which participates in the InFusion cloning process to generate plasmid pACG_araBAD.
ori2_InFusion_R	AGCGGATGCCGGGAGGTTTAAACAAGCAGGACA CAGCAGCAATCCACAG	
InFusion1_seq_F1	GAGCTGCATGTGTCTAGAGGTTTTTC	PCR-verification and sequencing of plasmid pACG_araBAD and pACG_XylSPm.
InFusion1_seq_F2	GCGACAAGCAAACATGCTGTG	
InFusion1_seq_F3	TCGGCAAACAAATTCTCGTCCCTG	
InFusion1_seq_F4	AAGATTAGCGGATCCTACCTGAC	
InFusion1_seq_F5	AAGCGGCTATTTAACGACCCTG	
InFusion1_seq_F6	ATCGTCGTGGTATTCACTCCAG	
InFusion1_seq_F7	ACCTCTACGTGCCGATCACTC	
InFusion1_seq_F8	AATAGCCCGCAATCGTCCAG	
InFusion1_seq_F9	AGTATCGTAAGCCGGATGGCTC	

InFusion1_seq_F10	TCTGAATTGGCTATCCGCGTG	
InFusion1_seq_F11	ATAACCCGGCGCTGGAGAATAG	
InFusion1_seq_F12	ATGCAATAAAGCCCACTTGCTG	
InFusion1_seq_F13	GGTAAAGGTCAGATCCGGATGAG	
InFusion1_seq_F14	TGGTCAACAGACACCGGCGTTC	
InFusion1_seq_F15	AACGTCATCTGCATCAAGAACTAG	
InFusion1_seq_F16	CTGGGCCCACTGTTCCACT	
InFusion1_seq_F17	GTCGATCAGACTATCAGCGTGAG	
InFusion1_seq_R1	GTTCCGTGCCGTTGTGAAG	
InFusion1_seq_R2	AGCGACCTGTTTGGGCAAATC	
InFusion1_seq_R3	TTCTTCTCTGAATGGCGGGAG	
InFusion1_seq_R4	ACCCCTCAAGACCCGTTTAGAG	
InFusion1_seq_R5	CAGGTTTCATCATGCCGTCTGTG	
InFusion1_seq_R6	CCTATAACCAGACCGTTCAGCTG	
InFusion1_seq_R7	ACACCTTCTCTAGAACCAGCATG	
InFusion1_seq_R8	ATCGCCGGCATCCTCTTCAG	
InFusion1_seq_R9	TGTGACGAACCACCCTCAAATC	
InFusion1_seq_R10	GCAACAACAAAATCGCAAAGTCATC	
InFusion1_seq_R11	CGATCACCGGTGGAAATACGTC	
InFusion1_seq_R12	AGTCAAACAACCTCAGCAGGCGTG	
InFusion1_seq_R13	AGCTGGTGTCGATAACGAAGTATC	
InFusion1_seq_R14	AATTCATTCTGCAATCGGCTTG	
InFusion1_seq_R15	ATTCAGGCCAGTTATGCTTTCTG	
InFusion1_seq_R16	AGACCGACAACACGAGTGGGATC	
<i>thrA</i> _NheI_F	AGCAGCTAGCATGCGAGTGTTGAAGTTCGGCGGT AC	PCR of native gene <i>thrA</i> from <i>E. coli</i> K-12 BW25113 genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>thrA</i> _NotI_R	ATTAGCGGCCGCCGACTCCTAACTCCATGAGAG GGTACGTAGCAGATC	
<i>ilvA</i> _NheI_F	AGCAGCTAGCATGGCTGACTCGCAACCCCTGTCC	PCR of native gene <i>ilvA</i> from <i>E.</i>

	GGTGCTC	<i>coli</i> K-12 BW25113 genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>ilvA</i> _NotI_R	ATTAGCGGCCGCCACCCGCCAAAAAGAACCTGAA CGCCGGGTTATTG	
<i>leuA</i> _NotI_F	ATTAGCGGCCGCCACGGTTTCCTTGTTGTTTCG TTG	PCR of native gene <i>leuA</i> from <i>E. coli</i> K-12 BW25113 genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>leuA</i> _NheI_R	AGCAGCTAGCATGAGCCAGCAAGTCATTATTTTC GATACCACATTG	
<i>ilvGM</i> _NotI_F	AGCAGCTAGCATGAATGGCGCACAGTGGGTGGT ACATGCGTTG	PCR of native operon <i>ilvGM</i> from <i>E. coli</i> K-12 BW25113 <i>ilvG</i> + genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>ilvGM</i> _NheI_R	ATTAGCGGCCGCCGGCGCGGATTTGTTGTGATGT GGTTGTGCTCTG	
<i>ilvIH</i> _NheI_F	AGCAGCTAGCATGGAGATGTTGTCTGGAGCCGAG ATGGTCGTC	PCR of native operon <i>ilvIH</i> from <i>E. coli</i> K-12 BW25113 genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>ilvIH</i> _NotI_R	ATTAGCGGCCGCCACGCATTATTTTATCGCCGCGC GAAAGTCCGAC	
<i>ilvC</i> _NheI_F	AGCAGCTAGCATGGCTAACTACTTCAATACACTGA ATCTGCGCCAG	PCR of native gene <i>ilvC</i> from <i>E. coli</i> K-12 BW25113 genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>ilvC</i> _NotI_R	ATTAGCGGCCGCCACCCGCAACAGCAATACGTTT CATATCTG	
<i>ilvBN</i> _NheI_F	ATTAGCGGCCGCCCTGAAAAAACACCGCATCTT GTAAACATC	PCR of native operon <i>ilvBN</i> from <i>E. coli</i> K-12 BW25113 genomic DNA for subsequent NheI- and NotI-mediated cloning into plasmid variants pACG_araBAD and pACG_XylSPm.
<i>ilvBN</i> _NotI_R	AGCAGCTAGCATGGCAAGTTCGGGCACAACATCG AC	
<i>leuA</i> _F1_seq	GGTGATCGAACGCGCTATCTATATG	
<i>leuA</i> _F2_seq	CAGCCAGTTAGTTAGCCAGATTTG	
<i>leuA</i> _F3_seq	GTCCGGTCGATGCCGTCTATC	
<i>leuA</i> _R1_seq	CGTCAACACCCATACGCTCAAG	
<i>leuA</i> _R2_seq	GGTATGTACGGAGATAATGGCTTTG	PCR-verification and sequencing of plasmid variants pACG_araBAD_LeuA and pACG_XylSPm_LeuA

<i>leuA</i> _R3_seq	CAGGAAAGCATCGTACAAATTGTC	
<i>ilvC</i> _F1_seq	CGTACAGCCACTGATGAAAGAC	
<i>ilvC</i> _F2_seq	GATGATGGACCGTCTCTTAACC	
<i>ilvC</i> _F3_seq	GCTGAAACCGTTTATGGCAGAG	PCR-verification and sequencing of plasmid variants pACG_araBAD_ <i>ilvC</i> and pACG_XylSPm_ <i>ilvC</i>
<i>ilvC</i> _R1_seq	CATGGTTTCGAACGCCAGTTC	
<i>ilvC</i> _R2_seq	TTCACCTCCGCAACGAAGGAC	
<i>ilvC</i> _R3_seq	TTCAGACCCTGTGCGCCACAG	
<i>ilvA</i> _F1_seq	ATCGTTATGCCAACCGCCAC	
<i>ilvA</i> _F2_seq	GTCGATAGCGATGCGATCTGTG	
<i>ilvA</i> _F3_seq	GCTACAGCGTGGTTGATC	PCR-verification and sequencing of plasmid variants pACG_araBAD_ <i>ilvA</i> and pACG_XylSPm_ <i>ilvA</i>
<i>ilvA</i> _R1_seq	GCACACCGACAAAGATGCAG	
<i>ilvA</i> _R2_seq	CATCCAGCGCTGCTTTCAG	
<i>ilvA</i> _R3_seq	GCTTTCTGTTCTTCCGTCAGG	
<i>ilvIH</i> _F1_seq	CTTTCAGGAGTGCACATGG	
<i>ilvIH</i> _F2_seq	GAATGCACGGTACCTACG	
<i>ilvIH</i> _F3_seq	CATTGACAAACCGCGTC	
<i>ilvIH</i> _F4_seq	GTCTACCCGATGCAGATTCG	
<i>ilvIH</i> _F5_seq	ATGTCACACCCTCGCTTTATAC	PCR-verification and sequencing of plasmid variants pACG_araBAD_ <i>ilvIH</i> and pACG_XylSPm_ <i>ilvIH</i>
<i>ilvIH</i> _R1_seq	CACGGTCTGGATGGTCATAC	
<i>ilvIH</i> _R2_seq	CGTATTGCAACGCGGTAGAC	
<i>ilvIH</i> _R3_seq	GGCGGATTCTTGCGACAAG	
<i>ilvIH</i> _R4_seq	GTGGGATTGTAAGAACGCATACTG	
<i>ilvIH</i> _R5_seq	GCCTCCGGGATAACCGAATAC	
<i>thrA</i> _F1_seq	CAGTGCCCGGATAGCATCAAC	
<i>thrA</i> _F2_seq	CCATCGCCCGAGTTCCAGATC	PCR-verification and sequencing of plasmid variants pACG_araBAD_ <i>thrA</i> and pACG_XylSPm_ <i>thrA</i>
<i>thrA</i> _F3_seq	GGATCTTCTGAACGCTCAATCTC	
<i>thrA</i> _F4_seq	CTCGTCGATGGATTACTACCATC	
<i>thrA</i> _F5_seq	ATTGATGAAGATGGCGTCTGC	

<i>thrA</i> _R1_seq	ATCAATAGTTTACGCGCCACATC	
<i>thrA</i> _R2_seq	CTCTTTGGCTTGCGCCAGTTC	
<i>thrA</i> _R3_seq	CGTTCAGCTCGCACACAGTC	
<i>thrA</i> _R4_seq	AGCAGAGTAGTCGGAACCGTTG	
<i>thrA</i> _R5_seq	TAAAGCATCCTGGCCGCTAATG	
<i>ilvGM</i> _F1_seq	CACCGTTTATCGGCACTGAC	
<i>ilvGM</i> _F2_seq	GCACCAAAGCGGCAAACTTC	
<i>ilvGM</i> _F3_seq	CTCCAGCGGTTTAGGTACCATG	
<i>ilvGM</i> _F4_seq	ATGTATCGGCTCGCTTCAATC	PCR-verification and sequencing of plasmid variants pACG_araBAD_ <i>ilvGM</i> and pACG_XylSPm_ <i>ilvGM</i>
<i>ilvGM</i> _R1_seq	GAAGGCGCTGGCTAACATGAG	
<i>ilvGM</i> _R2_seq	TCACCGGGATGGTCGTAAC	
<i>ilvGM</i> _R3_seq	ACCGCCAACGTACAGCATC	
<i>ilvGM</i> _R4_seq	GTAGCACGAGCATAACCGATAG	
<i>ilvBN</i> _R2_seq	GCCATAGCGGGCTGTGTTTC	PCR-verification and sequencing of plasmid pACG_XylSPm_ <i>ilvBN</i>
<i>ilvBN</i> _F1_seq	TGGACACCTACGGCATCTCTATC	
KO3_F	TTTACACATTTTTCCGTCAAACAGTGAGGCAGGC CGTGTAGGCTGGAGCTGCTTC	PCR from plasmids pKD3 or pKD4 to generate the <i>ilvIH</i> knock-out cassette.
KO3_R	ACATGTTGGGCTGTAAATTGCGCATTGAGATCATT CATGGGAATTAGCCATGGTCC	
<i>ilvIH</i> _F	TACCGCTGCCGTTAAAACAAAACAG	PCR-verification and sequencing of the target genomic <i>ilvIH</i> region.
<i>ilvIH</i> _R	CCAGTTTCACAATTGCCCTTGC	
<i>ilvIH</i> _F2	ATGGCAATGATGAGCGTACTTAGCGTGG	
<i>ilvA</i> _F	TTACAGGTAAGCGATGCCGAAGTGG	PCR-verification of the target genomic <i>ilvA</i> region
<i>ilvA</i> _R	GGTGCGTAATCAGGTGTCGGTAGAG	
<i>thrA</i> _F	GTTCTTACTGGTGTCCGATGGTGG	PCR-verification of the target genomic <i>thrA</i> region
<i>thrA</i> _R	CCAGCAAACGAGTGTCTTAAGCGG	
<i>leuA</i> _F	GGGCGCAGGTTGCTGAATAATTTG	PCR-verification of the target genomic <i>leuA</i> region
<i>leuA</i> _R	TTTACGGATGCAGAACTCACGCTG	
<i>ilvC</i> _F	GGATCATGCTAAAATCGCCGCTG	PCR-verification of the target

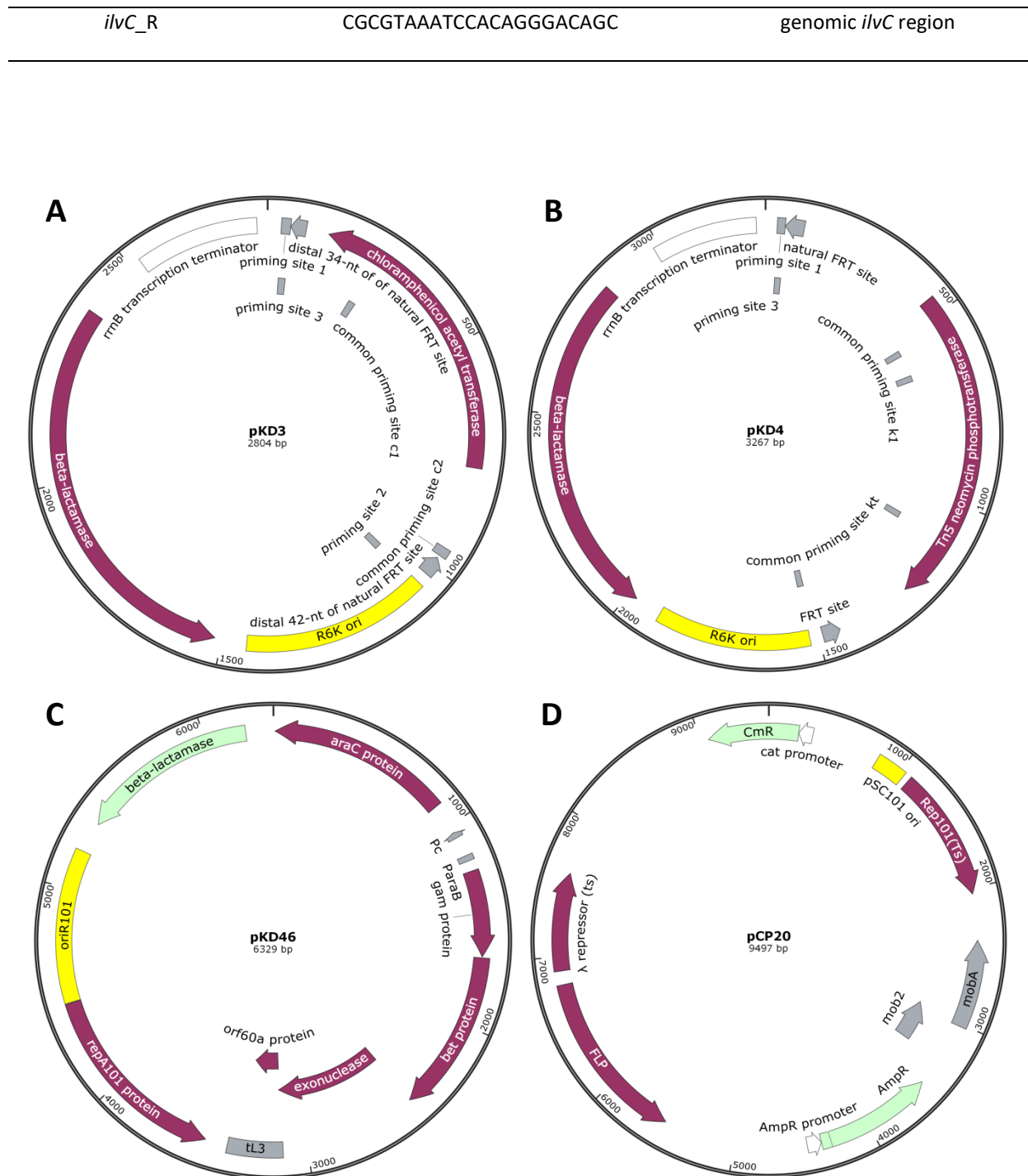


Figure S1. Plasmid maps of pKD3 (A), pKD4 (B), pKD46 (C) and pCP20 (D). The use of these plasmids allows homologous recombination in *E. coli*. Plasmid maps were generated by Snapgene®.

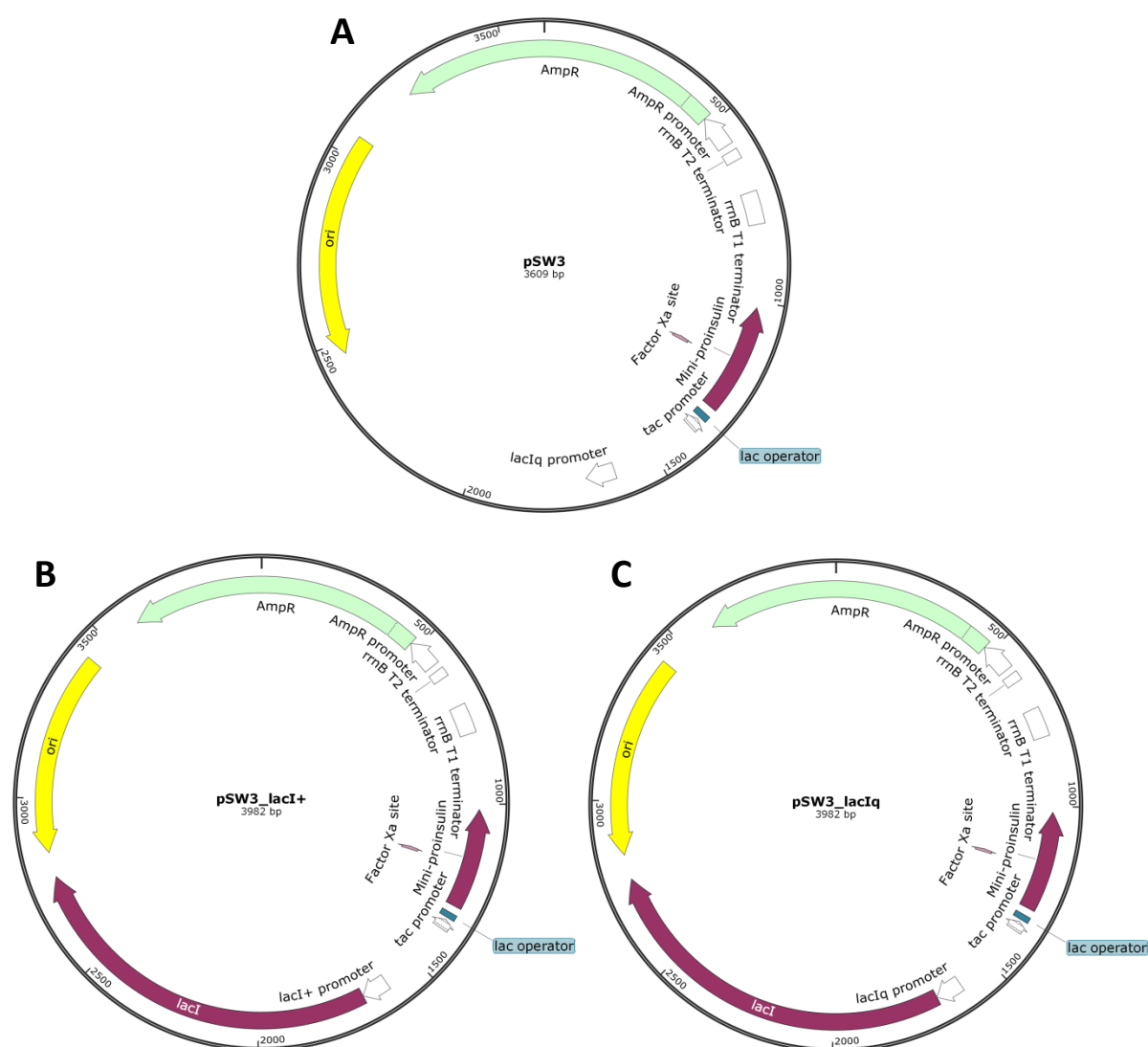


Figure S2. Plasmid maps of pSW3 (A), pSW3_lacI⁺ (B) and pSW3_lacI^q (C). Plasmid maps were generated by Snapgene®.

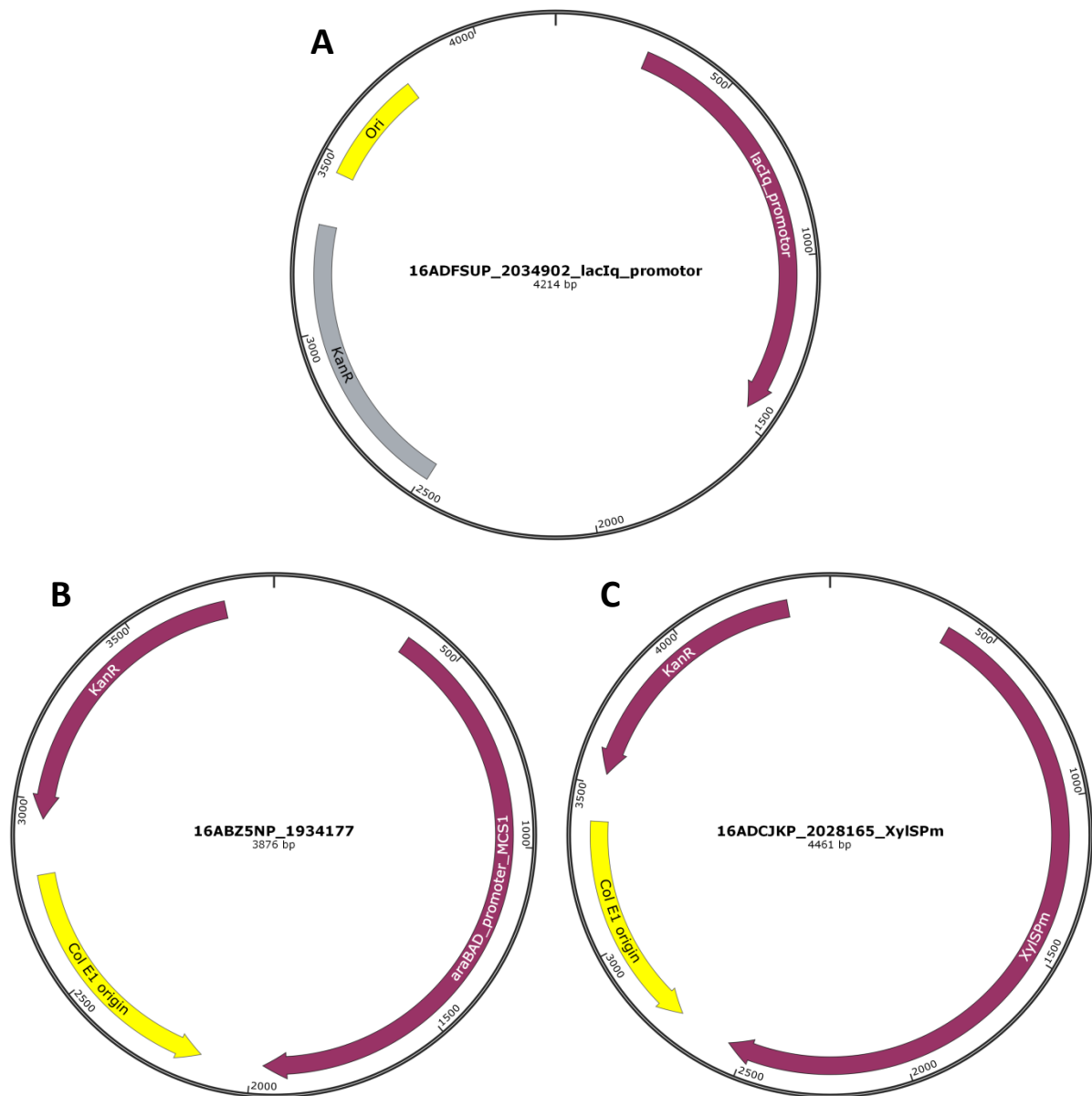


Figure S3. Plasmid maps of 16ADFSUP_2034902_lacIq_promotor (A), 16ABZ5NP_1934177 (B), 16ADCJKP_2028165_XylSPm (C). These plasmids were provided by the GeneArt® Gene Synthesis service of Termofisher scientific. Plasmid maps were generated by Snapgene®.

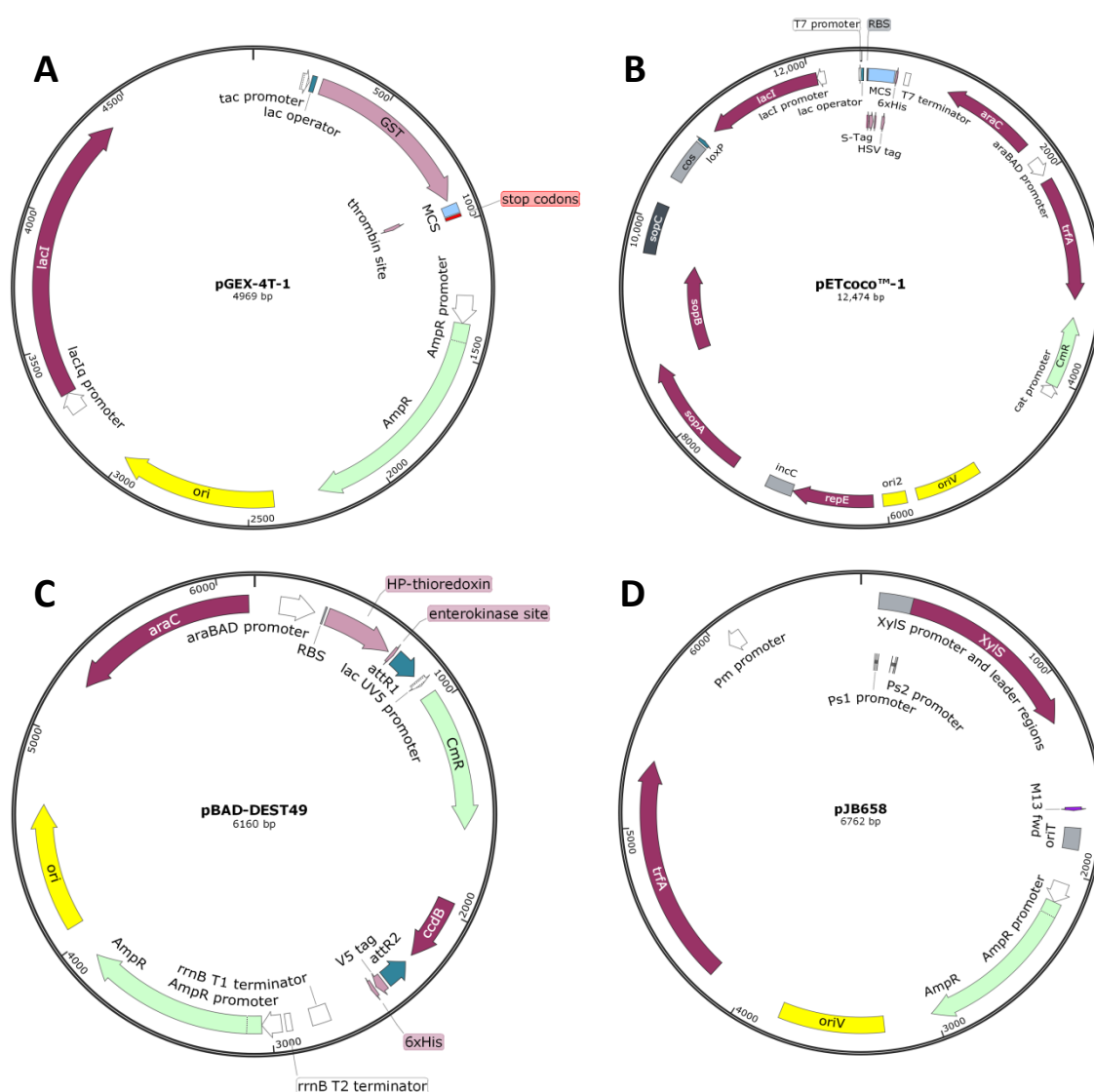
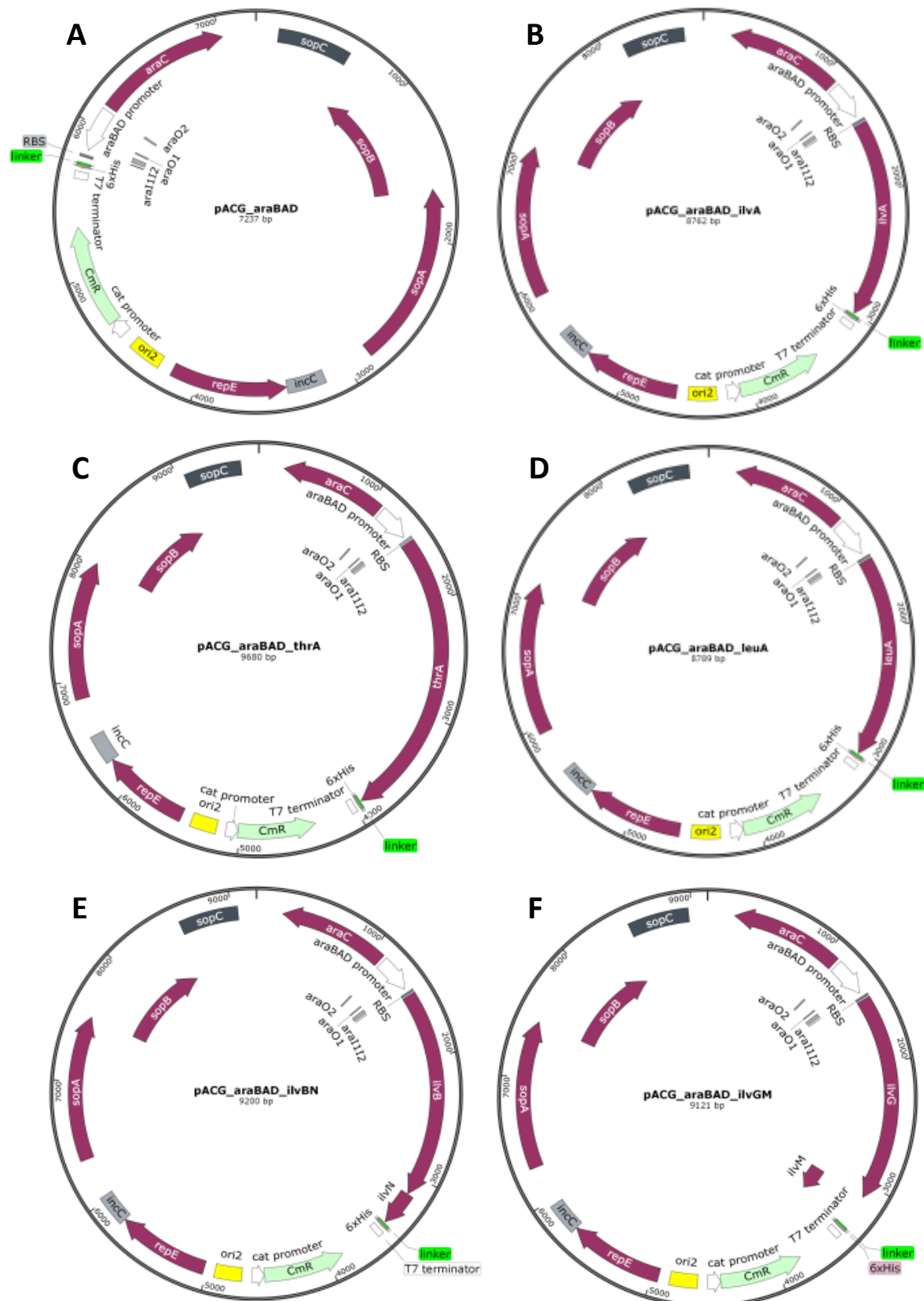


Figure S4. Plasmid map of pGEX-4T-1 (A), pETcoco-1 (B), pBAD_DEST49 (C) and pJB658 (D). Plasmid maps were generated by Snapgene®.



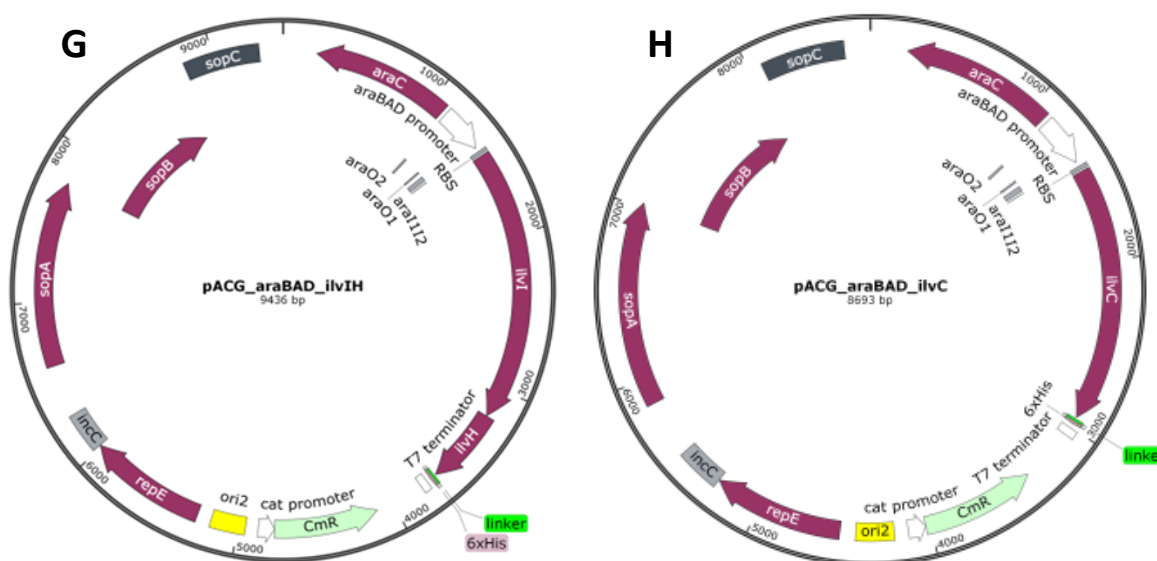
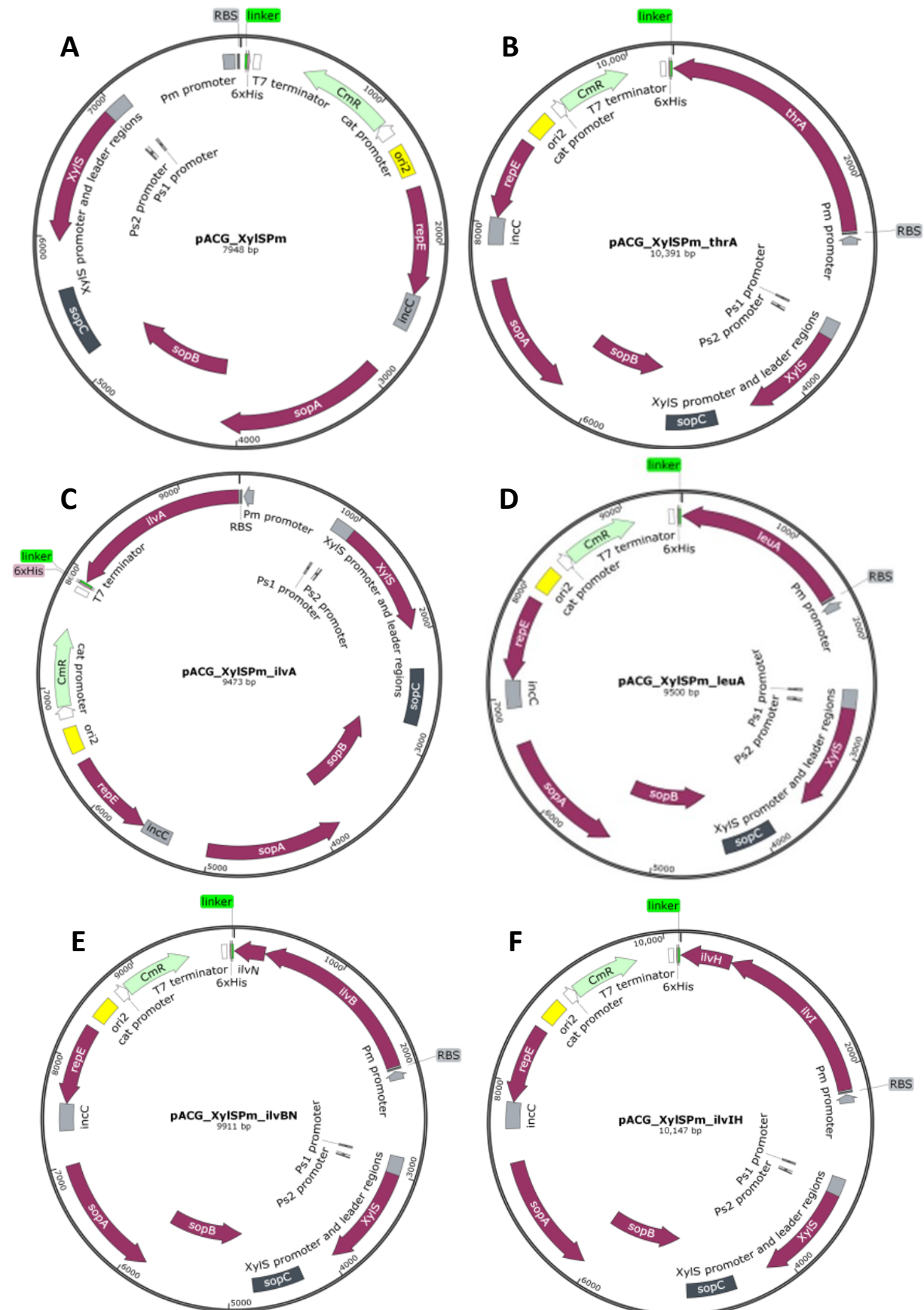


Figure S5. Plasmid map of the different pACG_araBAD plasmid variants generated in this study: pACG_araBAD (A), pACG_araBAD_ilvA (B), pACG_araBAD_thrA (C), pACG_araBAD_leuA (D), pACG_araBAD_ilvBN (E), pACG_araBAD_ilvGM (F), pACG_araBAD_ilvIH (G) and pACG_araBAD_ilvC (H). Plasmid maps were generated by Snapgene®.



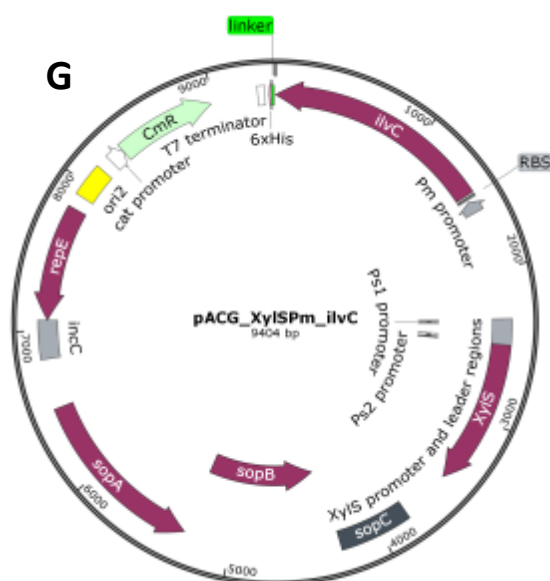


Figure S6. Plasmid map of the different pACG_XylSPm plasmid variants generated in this study: pACG_XylSPm (A), pACG_XylSPm_thrA (B), pACG_XylSPm_ilvA (C), pACG_XylSPm_leuA (D), pACG_XylSPm_ilvBN (E), pACG_XylSPm_ilvIH (F) and pACG_XylSPm_ilvC (G). Plasmid maps were generated by Snapgene®.

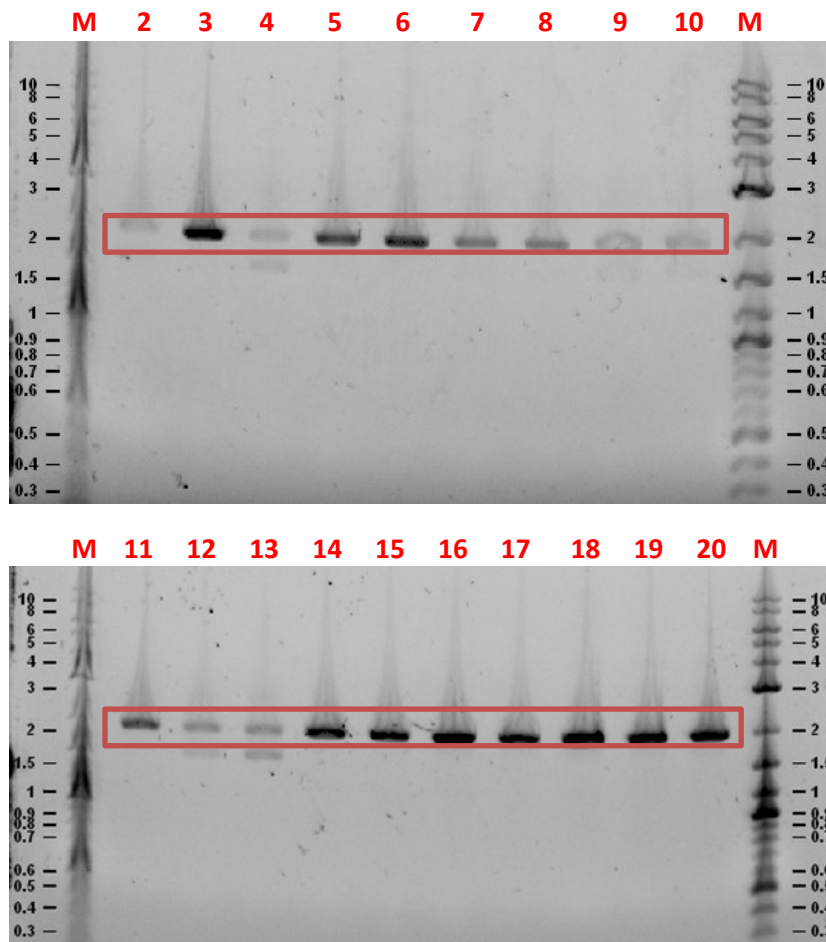


Figure S7. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pSW3_ *lacI*^q clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for *E. coli* K-12 NEB5α pSW3_ *lacI*^q is 2010 bp (indicated in red). 2-20: clones 2-20 potential *E. coli* K-12 NEB5α pSW3_ *lacI*^q; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

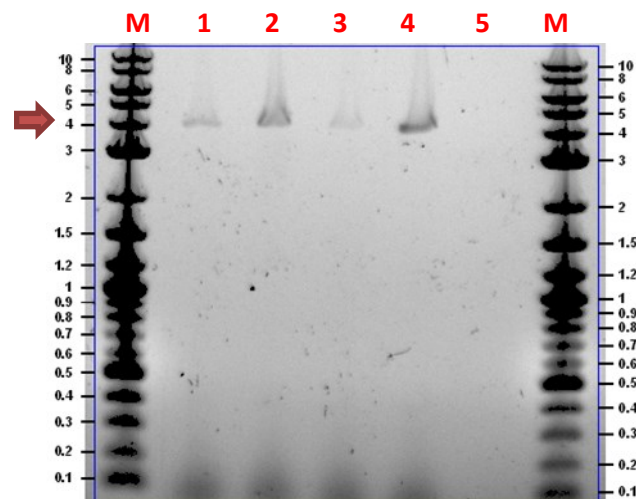


Figure S8. Gel electrophoresis of PCR products obtained by mutagenesis PCR of plasmid pSW3_ *lacI*^q by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size is 3982 bp (indicated by a red arrow). Four different buffer combinations were tested for PCR. 1: buffer HF; 2: buffer HF plus DMSO; 3: buffer GC; 4: buffer GC plus DMSO; 5: negative control; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

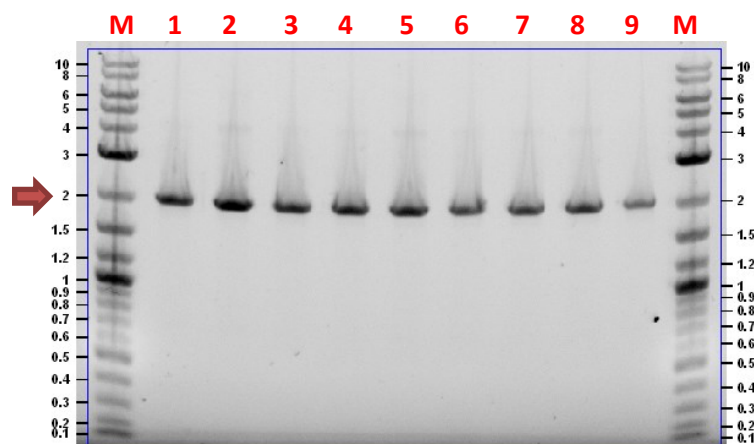


Figure S9. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pSW3_ *lacI*⁺ clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for *E. coli* K-12 NEB5α pSW3_ *lacI*⁺ is 2010 bp (indicated by a red arrow). 1-9: clones 1-9 potential *E. coli* K-12 NEB5α pSW3_ *lacI*⁺; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

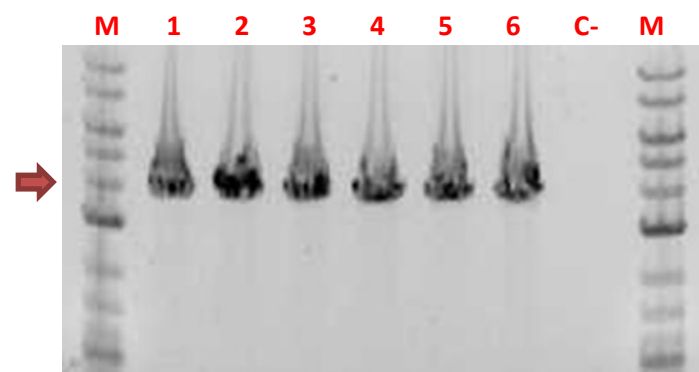


Figure S10. Gel electrophoresis of EcoRI-restriction products from plasmid preparations of potential *E. coli* K-12 BW25113 pSW3 clones for verification of plasmid pSW3 presence. Expected fragment size: 3609 bp (indicated by a red arrow). 1-6: clones 1-6 potential *E. coli* K-12 BW25113 pSW3; C-: negative control (no template was used for restriction); M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

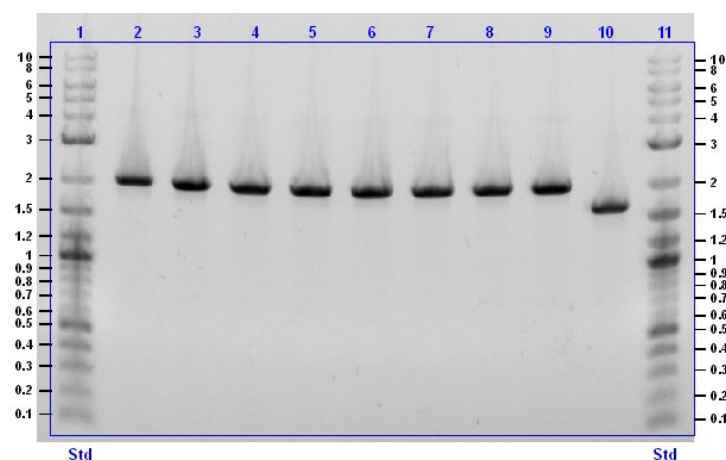


Figure S11. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 BW25113 pSW3_ *lacI*^q clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for *E. coli* K-12 BW25113 pSW3_ *lacI*^q is 2010 bp while for *E. coli* K-12 BW25113 pSW3 is 1637 bp. 2-9: clones 1-8 potential *E. coli* K-12 BW25113 pSW3_ *lacI*^q; 10: *E. coli* K-12 BW25113 pSW3; Std (1 and 11): molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

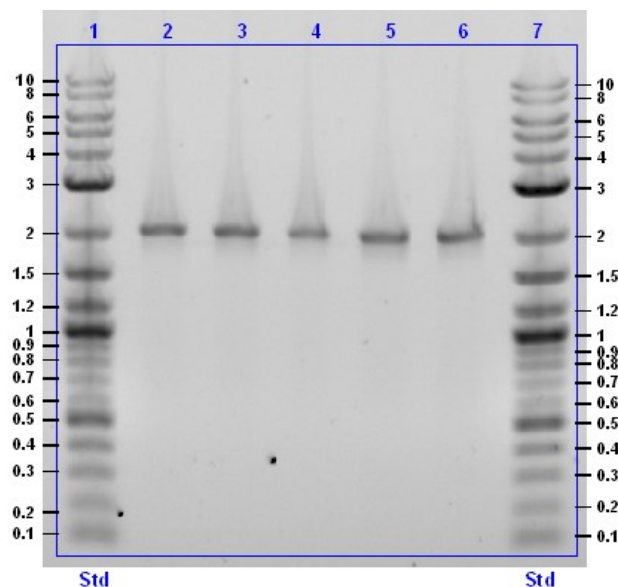


Figure S12. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ is 2010 bp. 2-6: clones 1-5 potential *E. coli* K-12 BW25113 pSW3_ *lacI*⁺; Std (1 and 7): molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

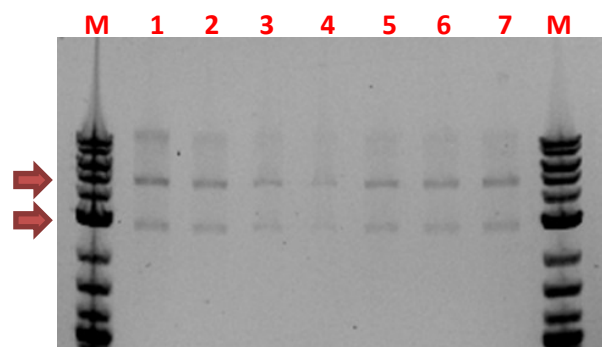


Figure S13. Gel electrophoresis of XhoI- and MssI-restriction products from plasmid preparations of potential *E. coli* K-12 NEB5α pACG_araBAD clones for verification of plasmid pACG_araBAD presence. Expected fragment sizes: 2678 and 4559 bp (indicated by red arrows). 1-7: clones 1-7 potential *E. coli* K-12 NEB5α pACG_araBAD; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

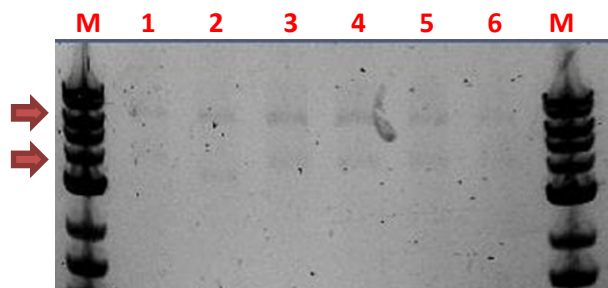


Figure S14. Gel electrophoresis of XhoI- and NheI-restriction products from plasmid preparations of potential *E. coli* K-12 NEB5α pACG_araBAD_ *thrA* clones for verification of plasmid pACG_araBAD_ *thrA* presence. Expected fragment sizes: 3658 and 6022 bp (indicated by red arrows). 1-6: clones 1-6 potential *E. coli* K-12 NEB5α pACG_araBAD_ *thrA*; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

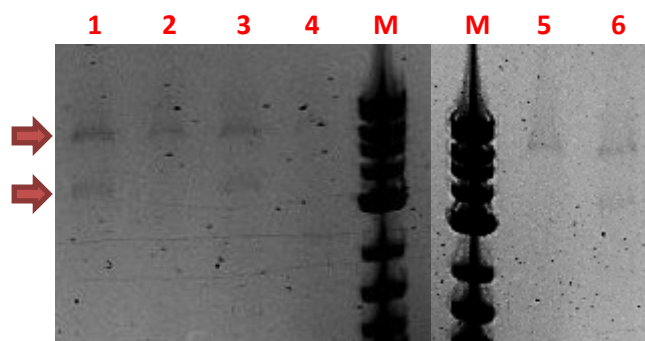


Figure S15. Gel electrophoresis of XhoI- and NheI-restriction products from plasmid preparations of potential *E. coli* K-12 NEB5α pACG_araBAD_ilmH clones for verification of plasmid pACG_araBAD_ilmH presence. Expected fragment sizes: 3414 and 6022 bp (indicated by red arrows). 1-6: clones 1-6 potential *E. coli* K-12 NEB5α pACG_araBAD_ilmH; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

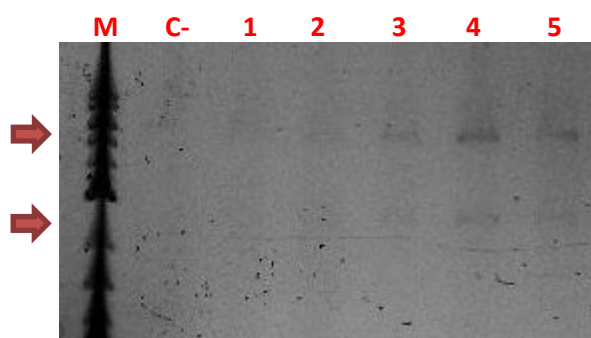


Figure S16. Gel electrophoresis of XhoI- and NheI-restriction products from plasmid preparations of potential *E. coli* K-12 NEB5α pACG_araBAD_ilmC clones for verification of plasmid pACG_araBAD_ilmC presence. Expected fragment sizes: 2671 and 6022 bp (indicated by red arrows). 1-5: clones 1-5 potential *E. coli* K-12 NEB5α pACG_araBAD_ilmC; C-: negative control (no template); M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

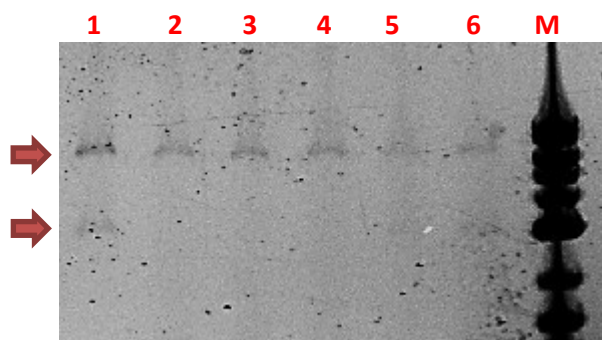


Figure S17. Gel electrophoresis of XhoI- and NheI-restriction products from plasmid preparations of potential *E. coli* K-12 NEB5α pACG_araBAD_ilmA clones for verification of plasmid pACG_araBAD_ilmA presence. Expected fragment sizes: 2740 and 6022 bp (indicated by red arrows). 1-6: clones 1-6 potential *E. coli* K-12 NEB5α pACG_araBAD_ilmA; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

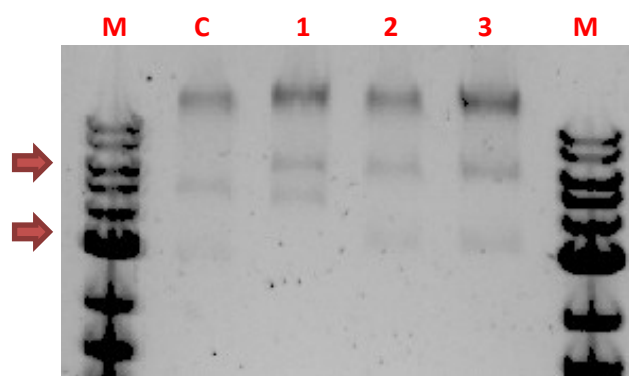


Figure S18. Gel electrophoresis of NotI- and MspI-restriction products from plasmid preparations of potential *E. coli* K-12 NEB5 α pACG_araBAD_LeuA clones for verification of plasmid pACG_araBAD_LeuA presence. Expected fragment sizes: 3040 and 5749 bp (indicated by red arrows). 1-3: clones 1-3 potential *E. coli* K-12 NEB5 α pACG_araBAD_LeuA; C: plasmid pACG_araBAD; M: molecular weight marker (Quick-Load[®] 2-Log DNA Ladder (0.1-10.0 kb)).

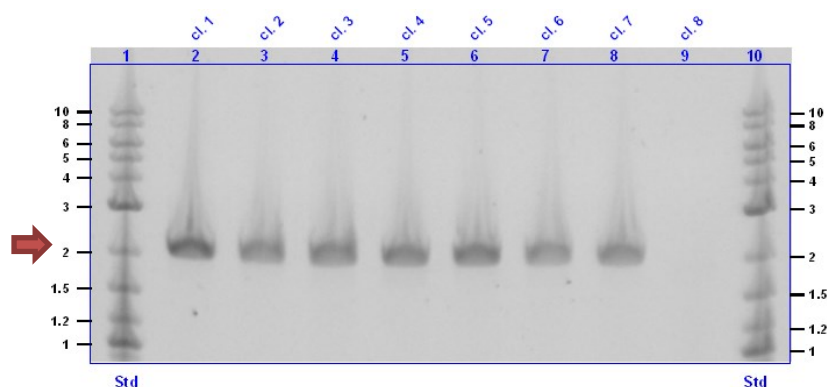


Figure S19. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5 α pACG_araBAD_LeuGM clones by using primers InFusion1_seq_F4 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5 α pACG_araBAD_LeuGM is 2127 bp (indicated by a red arrow). 2-9: clones 1-8 potential *E. coli* K-12 NEB5 α pACG_araBAD_LeuGM; Std (1 and 10): molecular weight marker (Quick-Load[®] 2-Log DNA Ladder (0.1-10.0 kb)). Molecular weights of the DNA standard are shown in kb.

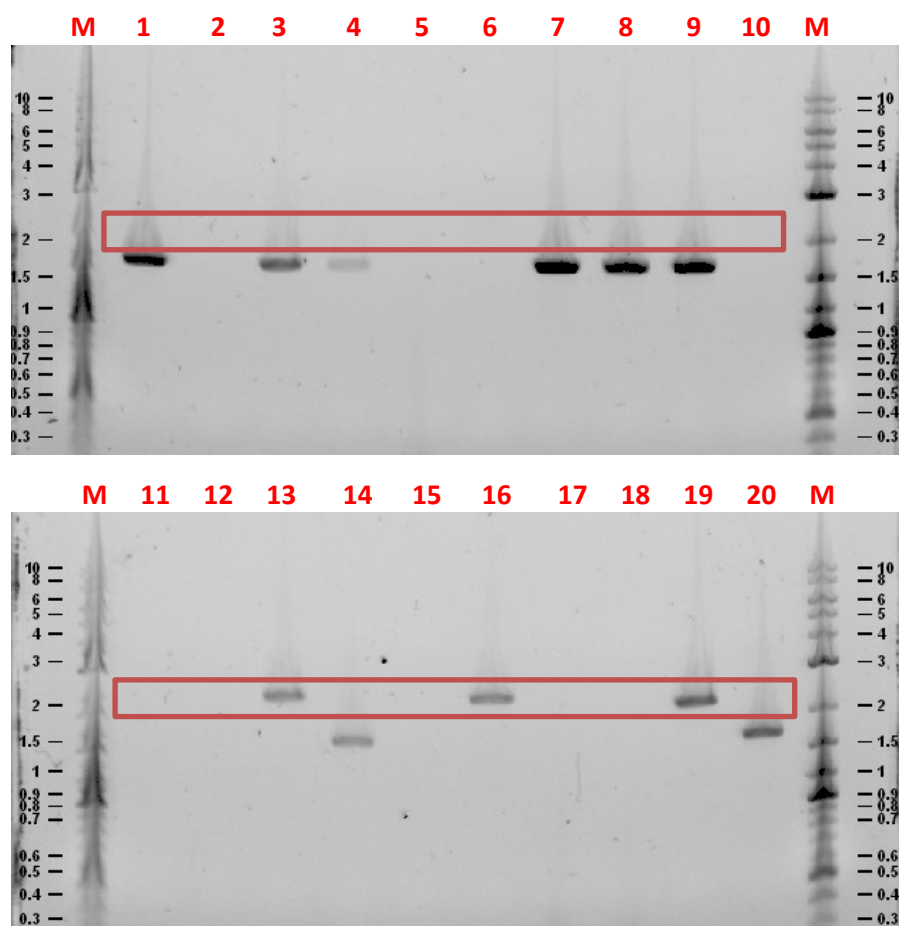


Figure S20. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm is 2453 bp (indicated by red rectangles). 1-20: clones 1-20 potential *E. coli* K-12 NEB5α pACG_XylSPm; Std (1 and 10): M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb). Molecular weights of the DNA standard are shown in kb.

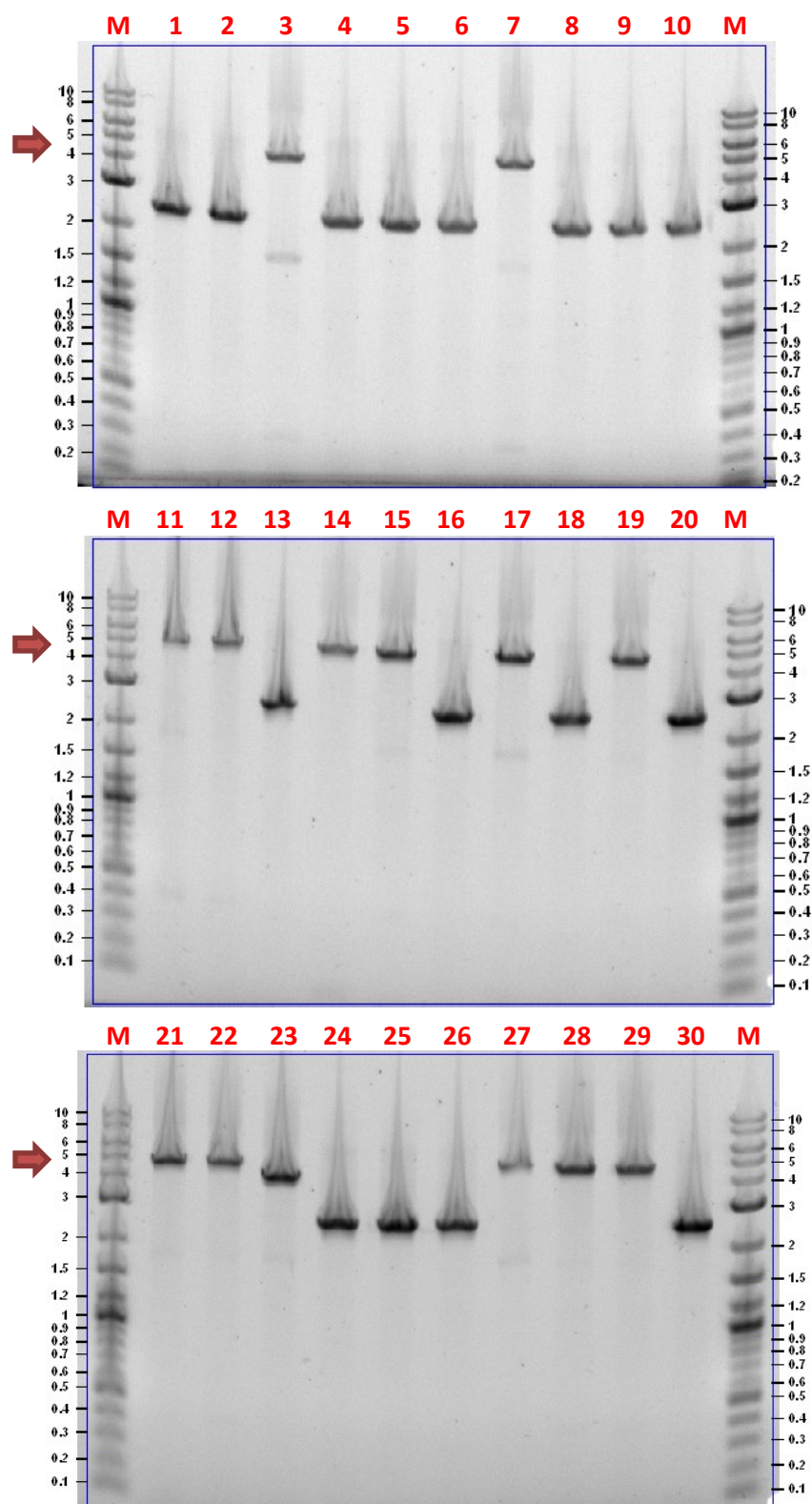


Figure S21. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm_ilmH clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm_ilmH is 4652 bp (indicated by the red arrow). 1-30: clones 1-30 potential *E. coli* K-12 NEB5α pACG_XylSPm_ilmH; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb). Molecular weights of the DNA standard are shown in kb.

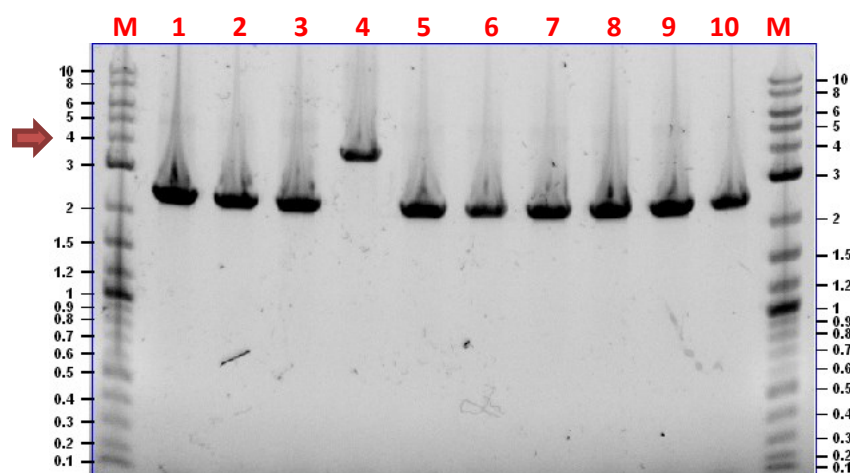


Figure S22. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm_LeuA clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm_LeuA is 4005 bp (indicated by the red arrow). 1-10: clones 1-10 potential *E. coli* K-12 NEB5α pACG_XylSPm_LeuA; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb). Molecular weights of the DNA standard are shown in kb.

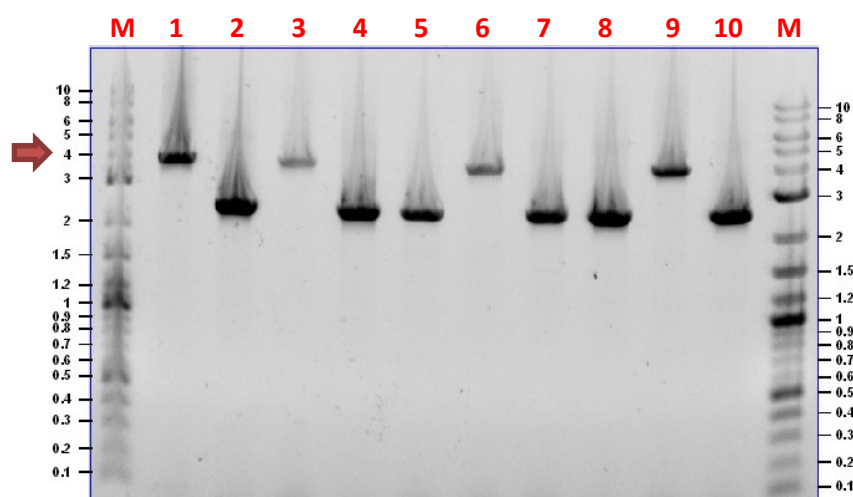


Figure S23. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm_ilvC clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm_ilvC is 3909 bp (indicated by the red arrow). 1-10: clones 1-10 potential *E. coli* K-12 NEB5α pACG_XylSPm_ilvC; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb). Molecular weights of the DNA standard are shown in kb.

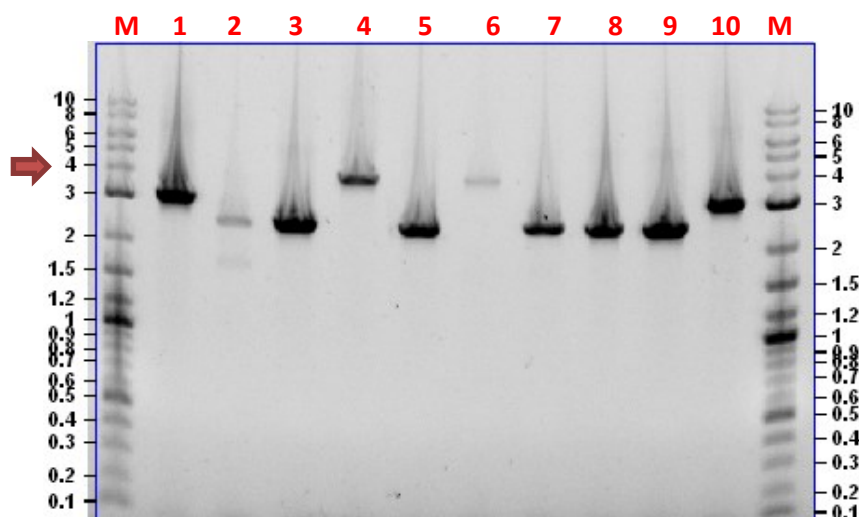


Figure S24. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm_*ilvA* clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm_*ilvA* is 3978 bp (indicated by the red arrow). 1-10: clones 1-10 potential *E. coli* K-12 NEB5α pACG_XylSPm_*ilvA*; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb). Molecular weights of the DNA standard are shown in kb.

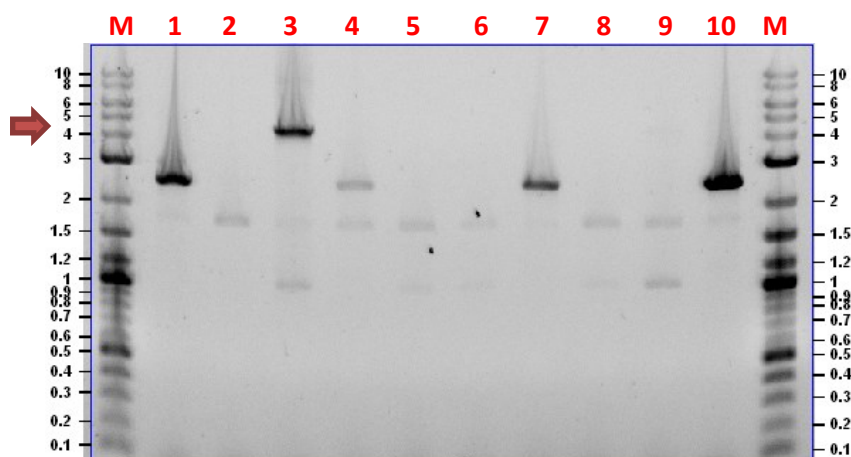


Figure S25. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm_*ilvBN* clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm_*ilvBN* is 4416 bp (indicated by the red arrow). 1-10: clones 1-10 potential *E. coli* K-12 NEB5α pACG_XylSPm_*ilvBN*; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb). Molecular weights of the DNA standard are shown in kb.

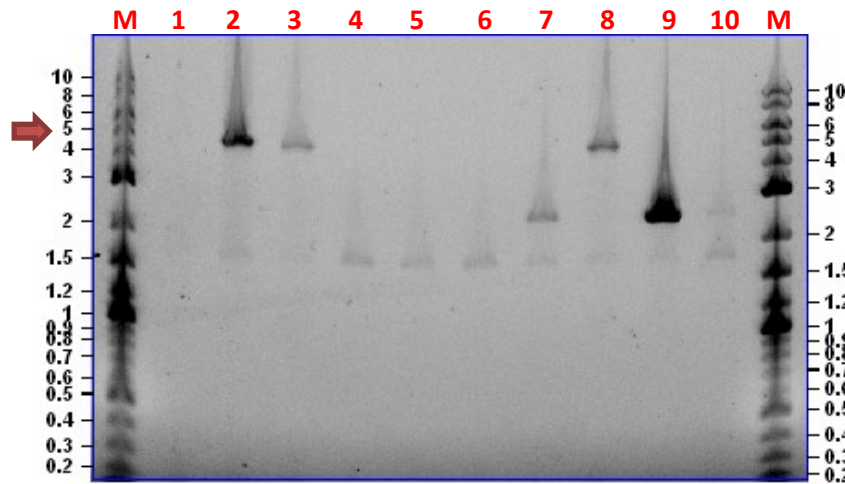


Figure S26. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 NEB5α pACG_XylSPm_thrA clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for *E. coli* K-12 NEB5α pACG_XylSPm_thrA is 4896 bp (indicated by the red arrow). 1-10: clones 1-10 potential *E. coli* K-12 NEB5α pACG_XylSPm_thrA; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)). Molecular weights of the DNA standard are shown in kb.

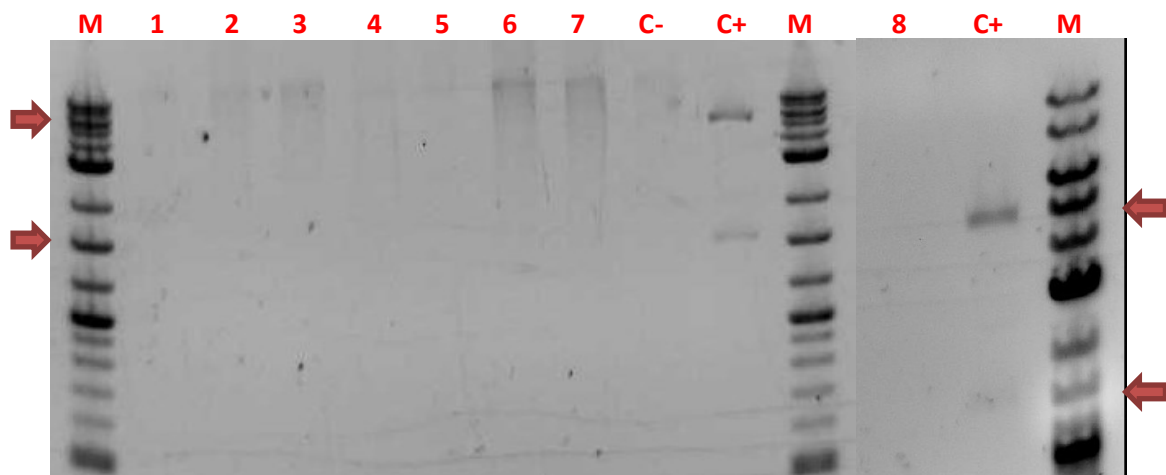


Figure S27. Gel electrophoresis of EcoRI-restriction products from plasmid preparations of potential *E. coli* K-12 BW25113 *geneX:kanR* mutants for verification of plasmid pKD46 curation. Expected fragment size by pKD46 presence: 1509 and 4820 bp (indicated by red arrows). In case of pKD46 absence, no band is expected. 1: *E. coli* K-12 BW25113 *leuA:kanR*; 2: *E. coli* K-12 BW25113 *ilvA:kanR*; 3: *E. coli* K-12 BW25113 *ilvB:kanR*; 4: *E. coli* K-12 BW25113 *ilvC:kanR*; 5: *E. coli* K-12 BW25113 *thrA:kanR*; 6: *E. coli* K-12 BW25113 *thrB:kanR*; 7: *E. coli* K-12 BW25113 *thrC:kanR*; 8: *E. coli* K-12 BW25113 *ilvI:kanR*; C-: negative control (*E. coli* K-12 BW25113); C+: positive control (pKD46); M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

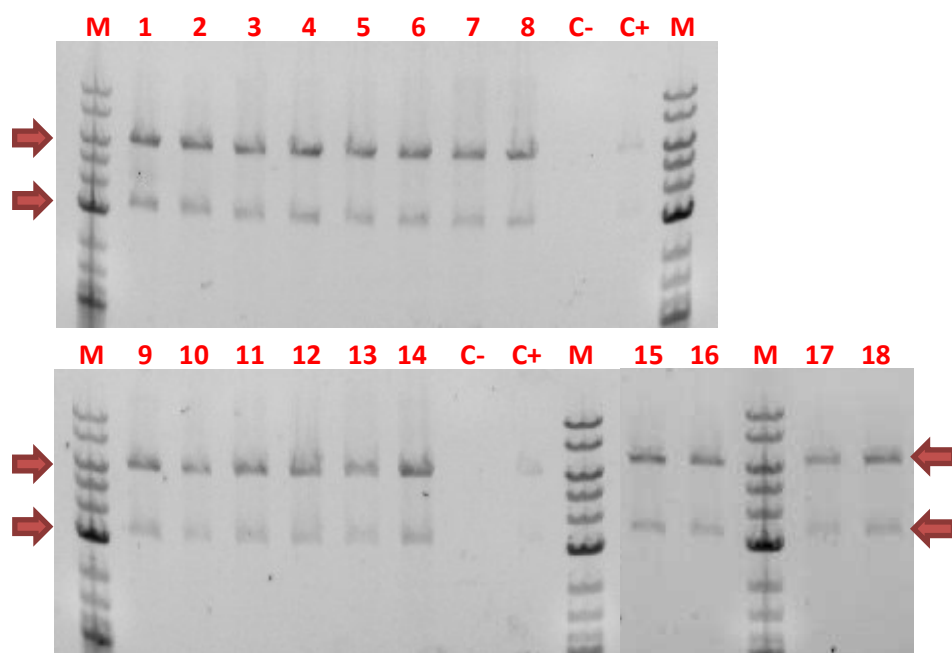


Figure S28. Gel electrophoresis of EcoRI-restriction products from plasmid preparations of potential *E. coli* K-12 BW25113 *geneX:kanR* pCP20 mutants for verification of plasmid pCP20 presence. Expected fragment size by pCP20 presence: 3454 and 6043 bp (indicated by red arrows). In case of pCP20 absence, no band is expected. 1-2: *E. coli* K-12 BW25113 *thrA:kanR* pCP20 clones 1-2; 3-4: *E. coli* K-12 BW25113 *thrB:kanR* pCP20 clones 1-2; 5-6: *E. coli* K-12 BW25113 *thrC:kanR* pCP20 clones 1-2; 7-8: *E. coli* K-12 BW25113 *leuA:kanR* pCP20 clones 1-2; 9-10: *E. coli* K-12 BW25113 *ilvC:kanR* pCP20 clones 1-2; 11-12: *E. coli* K-12 BW25113 *ilvA:kanR* pCP20 clones 1-2; 13-14: *E. coli* K-12 BW25113 *ilvB:kanR* pCP20 clones 1-2; 15-18: *E. coli* K-12 BW25113 *ilvI:kanR* pCP20 clones 1-4; C+: positive control (pCP20); C-: negative control (*E. coli* K-12 BW25113); M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

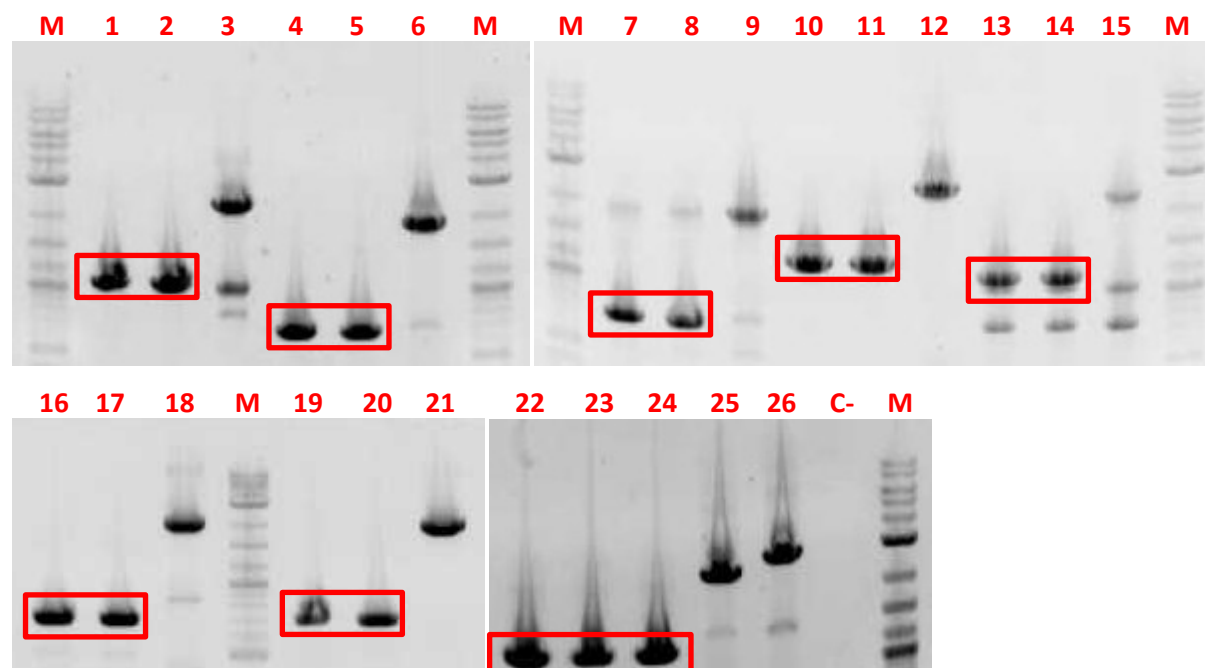


Figure S29. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 BW25113 Δ geneX pCP20 clones by using primers mentioned in Table 22 to verify removal of the kanamycin resistance cassette from the genome. The expected PCR product size for each mutant is shown in Table 22 (indicated with red squares for kanS variants). 1-2: *E. coli* K-12 BW25113 Δ thrB pCP20 (kanS) clones 1-2; 3: *E. coli* K-12 BW25113 thrB:kanR pCP20 (kanR); 4-5: *E. coli* K-12 BW25113 Δ thrC pCP20 (kanS) clones 1-2; 6: *E. coli* K-12 BW25113 thrC:kanR pCP20 (kanR); 7-8: *E. coli* K-12 BW25113 Δ ilvA pCP20 (kanS) clones 1-2; 9: *E. coli* K-12 BW25113 ilvA:kanR pCP20 (kanR); 10-11: *E. coli* K-12 BW25113 Δ ilvB pCP20 (kanS) clones 1-2; 12: *E. coli* K-12 BW25113 ilvB:kanR pCP20 (kanR); 13-14: *E. coli* K-12 BW25113 Δ thrA pCP20 (kanS) clones 1-2; 15: *E. coli* K-12 BW25113 thrA:kanR pCP20 (kanR); 16-17: *E. coli* K-12 BW25113 Δ leuA pCP20 (kanS) clones 1-2; 18: *E. coli* K-12 BW25113 leuA:kanR pCP20 (kanR); 19-20: *E. coli* K-12 BW25113 Δ ilvC pCP20 (kanS) clones 1-2; 21: *E. coli* K-12 BW25113 ilvC:kanR pCP20 (kanR); 22-24: *E. coli* K-12 BW25113 Δ ilvI pCP20 (kanS) clones 1-3; 25: *E. coli* K-12 BW25113 ilvI:kanR pCP20 (kanR); 26: *E. coli* K-12 BW25113; C-: no DNA; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

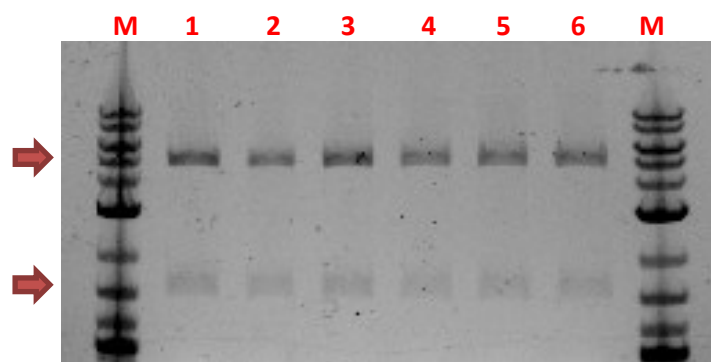


Figure S30. Gel electrophoresis of EcoRI-restriction products from plasmid preparations of potential *E. coli* K-12 BW25113 pKD46 clones for verification of plasmid pKD46 presence. Expected fragment sizes: 1509 and 4820 bp (indicated by red arrows). 1-6: clones 1-6 potential *E. coli* K-12 BW25113 pKD46. M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

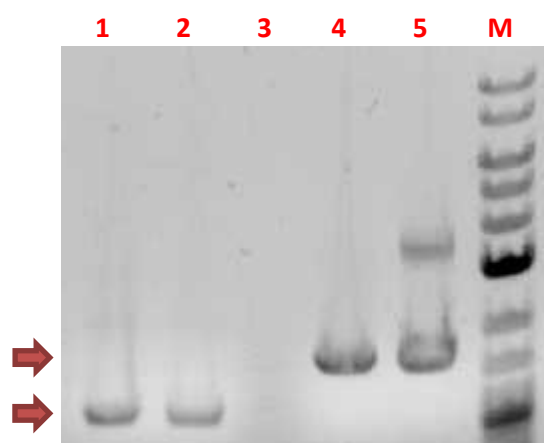


Figure S31. Gel electrophoresis of the *ilvIH* knock-out cassette obtained by PCR under different templates and buffer conditions. Expected fragment sizes (indicated by red arrows): 1105 bp (pKD3) or 1568 bp (pKD4). 1: template pKD3, buffer HF plus DMSO; 2: template pKD3, buffer GC plus DMSO; 3: negative control; 4: template pKD4, buffer HF plus DMSO; 5: template pKD4, buffer GC plus DMSO; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

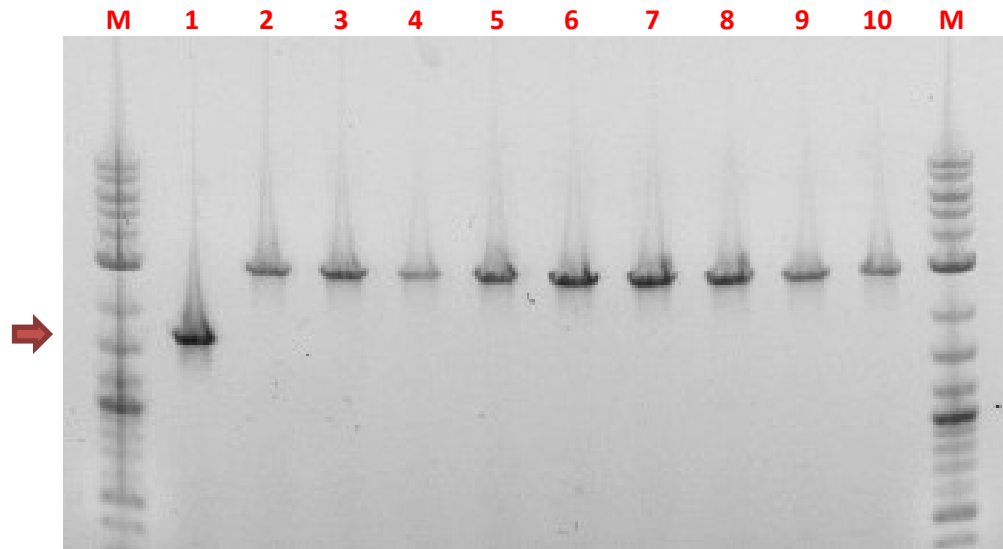


Figure S32. Gel electrophoresis of PCR products obtained from genomic DNA of potential *E. coli* K-12 BW25113 *ilvIH:cmR* pKD46 clones by using primers *ilvIH_F* and *ilvIH_R*. The expected PCR product size for *ilvIH* mutants is 1658 bp (indicated by a red arrow) while for wild type is 2844 bp. 1-10: clones 1-10 potential *E. coli* K-12 BW25113 *ilvIH:cmR* pKD46; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

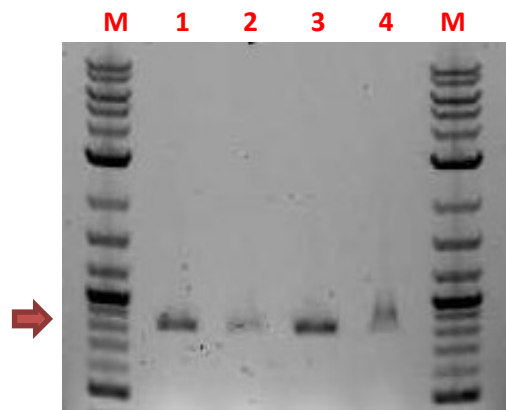


Figure S33. Gel electrophoresis of PCR products obtained from genomic DNA of potential *E. coli* K-12 BW25113 $\Delta ilvIH$ pCP20 clones by using primers *ilvIH_F2* and *ilvIH_R*. The expected PCR product size for *E. coli* K-12 BW25113 $\Delta ilvIH$ pCP20 is 834bp (indicated by a red arrow). 1-4: clones 1-4 potential *E. coli* K-12 BW25113 $\Delta ilvIH$ pCP20; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

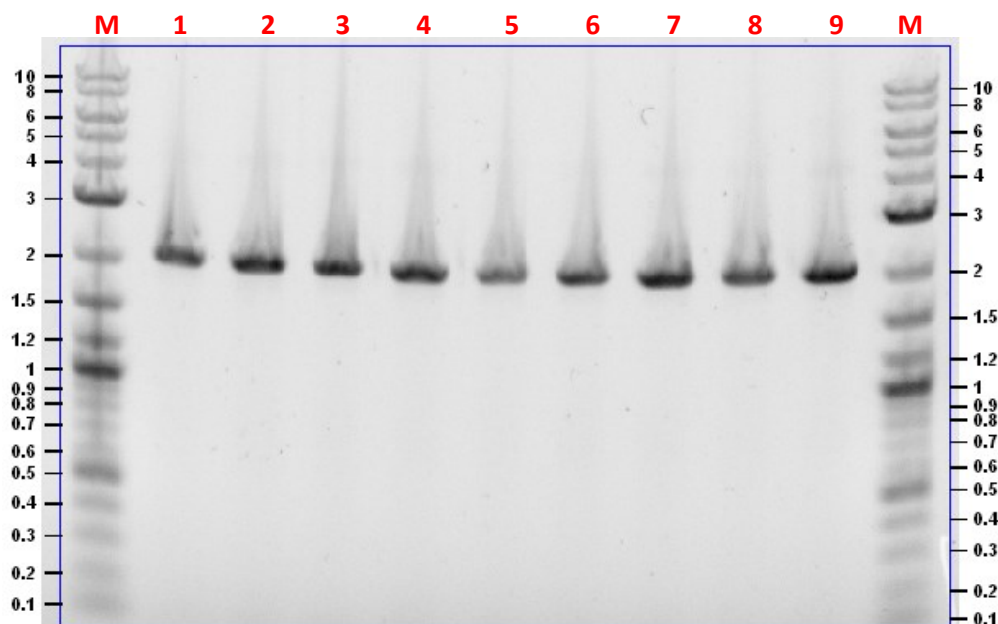


Figure S34. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for positive clones is 2010 bp. 1-3: *E. coli* K-12 BW25113 $\Delta ilvA$ pSW3_ *lacI*⁺ clones 1-3; 4-6: *E. coli* K-12 BW25113 $\Delta ilvBN$ pSW3_ *lacI*⁺ clones 1-3; 7-9: *E. coli* K-12 BW25113 $\Delta ilvC$ pSW3_ *lacI*⁺ clones 1-3; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

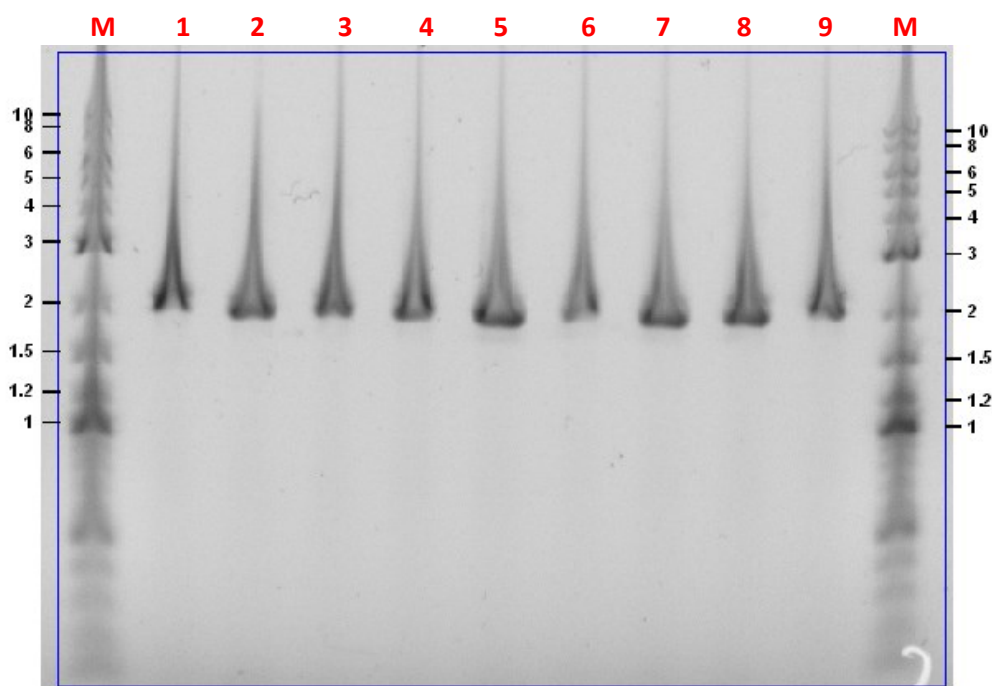


Figure S35. Gel electrophoresis of PCR products obtained by colony PCR of potential *E. coli* K-12 BW25113 $\Delta geneX$ pSW3_ *lacI*⁺ clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for positive clones is 2010 bp. 1-5: *E. coli* K-12 BW25113 $\Delta leuA$ pSW3_ *lacI*⁺ clones 1-5; 6-9: *E. coli* K-12 BW25113 $\Delta thrA$ pSW3_ *lacI*⁺ clones 1-4; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

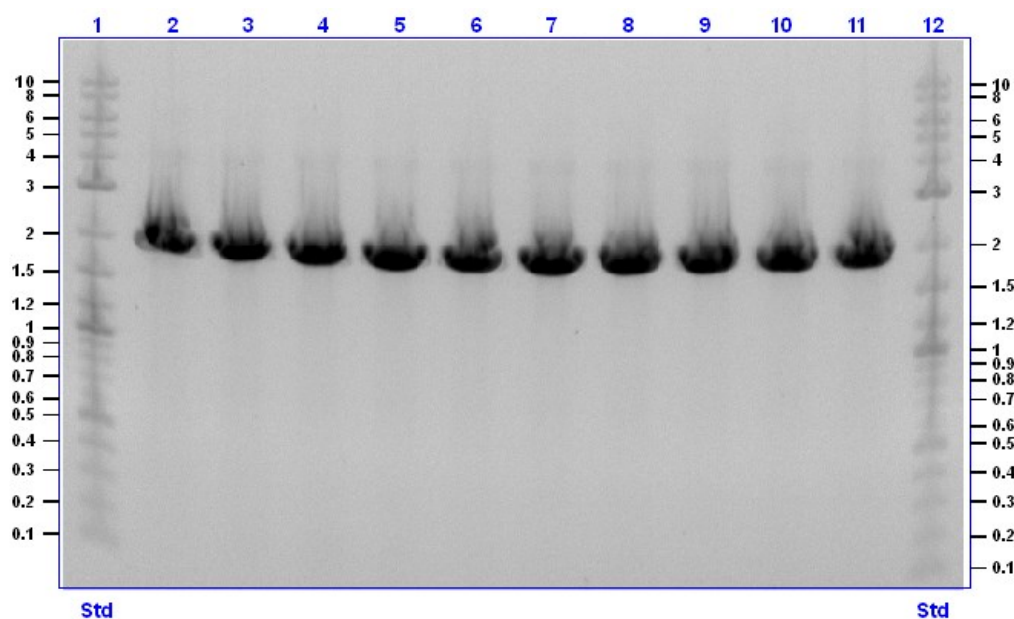


Figure S36. Gel electrophoresis of PCR products obtained by colony PCR of 10 potential *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for positive clones is 2010 bp. 2-11: *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ clones 1-10; 1, 12: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

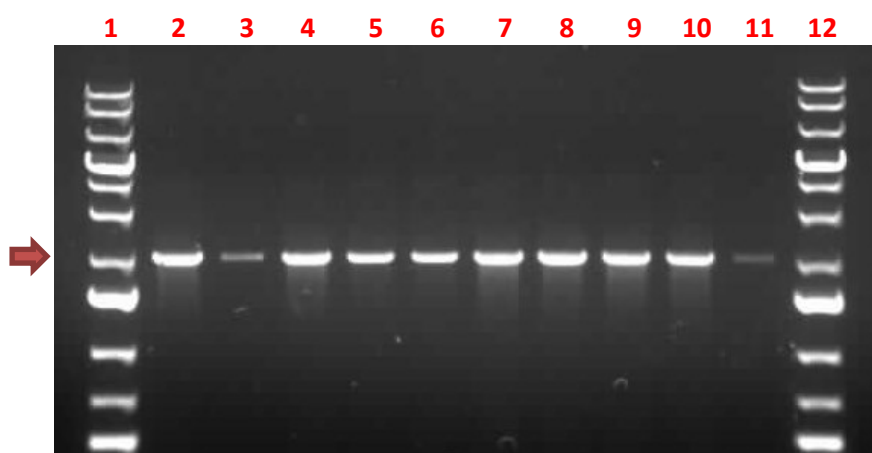


Figure S37. Gel electrophoresis of PCR products obtained by colony PCR of 10 potential *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_ *lacI*⁺ clones by using primers pSW3_R3_seq and pSW3_R6_seq. The expected PCR product size for positive clones is 2010 bp (indicated by a red arrow). 2-11: *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_ *lacI*⁺ clones 1-10; 1, 12: molecular weight marker (GeneRuler™ 1 kb Plus DNA Ladder).

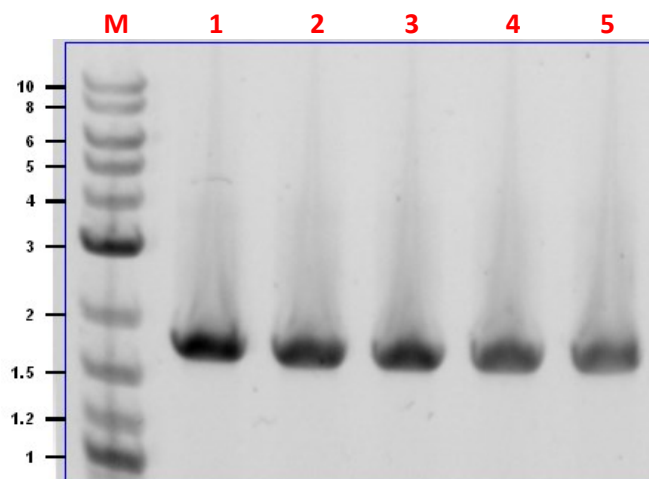


Figure S38. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta leuA$ pSW3_ *lacI*⁺ pACG_araBAD_ *leuA* clones by using primers InFusion1_seq_F4 an InFusion1_seq_R4. The expected PCR product size for positive clones is 1795 bp. 1-5: *E. coli* K-12 BW25113 $\Delta leuA$ pSW3_ *lacI*⁺ pACG_araBAD_ *leuA* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

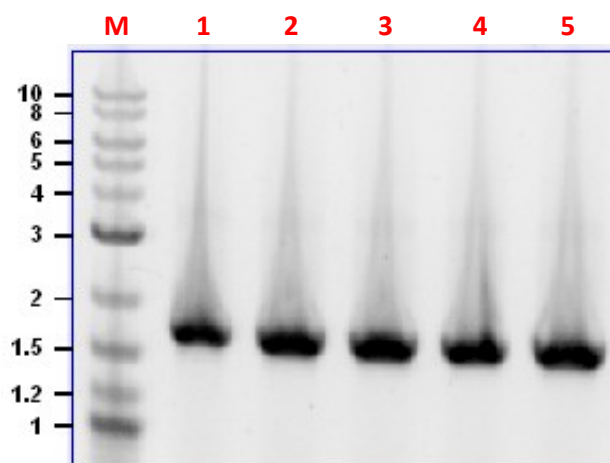


Figure S39. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvC$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvC* clones by using primers InFusion1_seq_F4 an InFusion1_seq_R4. The expected PCR product size for positive clones is 1699 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvC$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvC* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

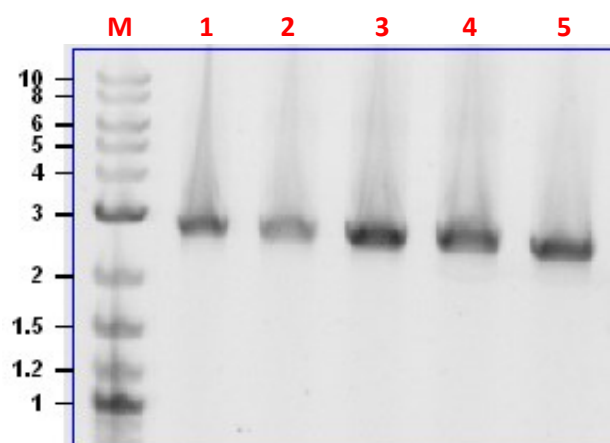


Figure S40. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta thrA$ pSW3_ *lacI*⁺ pACG_arabAD_ *thrA* clones by using primers InFusion1_seq_F4 an InFusion1_seq_R4. The expected PCR product size for positive clones is 2686 bp. 1-5: *E. coli* K-12 BW25113 $\Delta thrA$ pSW3_ *lacI*⁺ pACG_arabAD_ *thrA* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

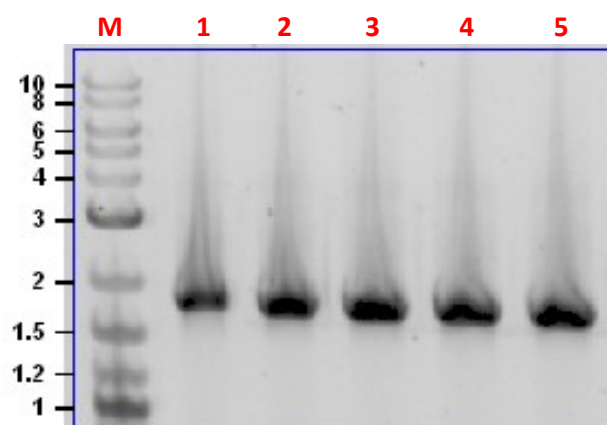


Figure S41. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvA$ pSW3_ *lacI*⁺ pACG_arabAD_ *ilvA* clones by using primers InFusion1_seq_F4 an InFusion1_seq_R4. The expected PCR product size for positive clones is 1768 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvA$ pSW3_ *lacI*⁺ pACG_arabAD_ *ilvA* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

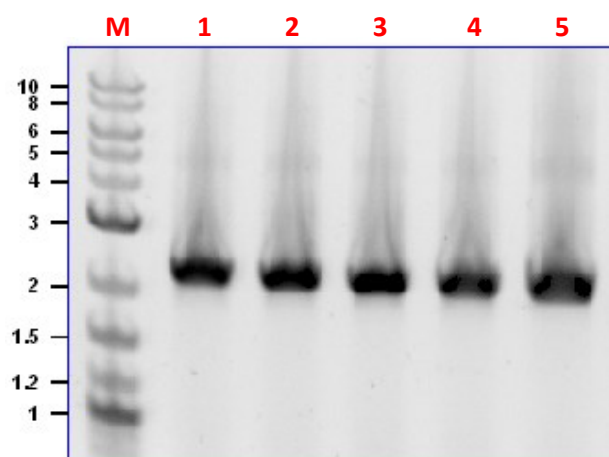


Figure S42. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvBN$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvBN* clones by using primers InFusion1_seq_F4 and InFusion1_seq_R4. The expected PCR product size for positive clones is 2206 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvBN$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvBN* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

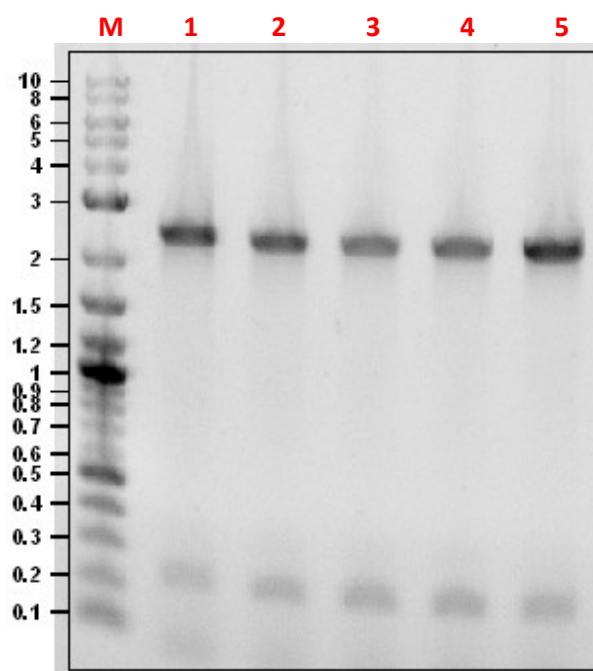


Figure S43. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvIH$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH* clones by using primers InFusion1_seq_F4 and InFusion1_seq_R4. The expected PCR product size for positive clones is 2442 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvIH$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

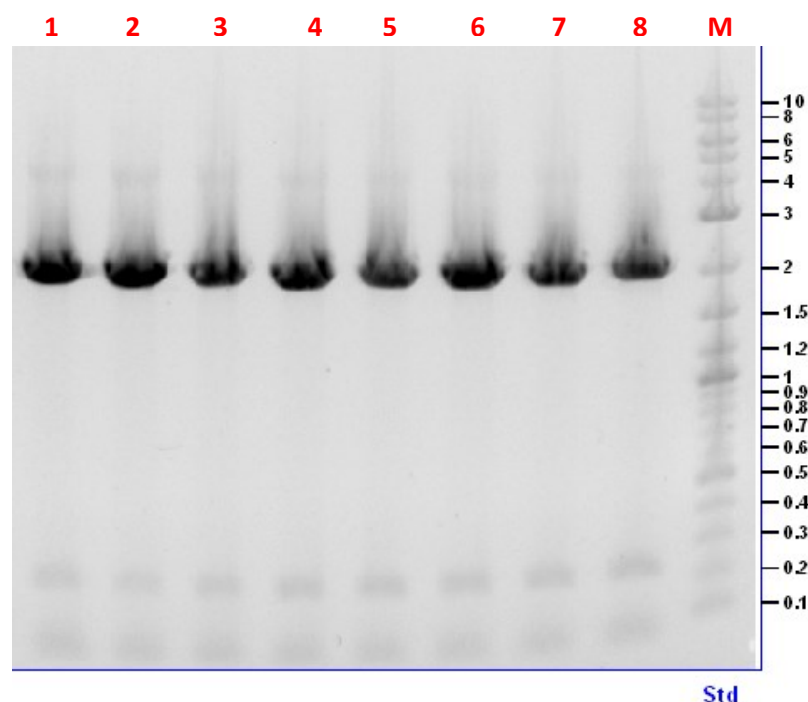


Figure S44. Gel electrophoresis of PCR products obtained by colony PCR of 8 potential *E. coli* K-12 BW25113 $\Delta ilvGM$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* clones by using primers InFusion1_seq_F4 an InFusion1_seq_R4. The expected PCR product size for positive clones is 2127 bp. 1-8: *E. coli* K-12 BW25113 $\Delta ilvGM$ pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* clones 1-8; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

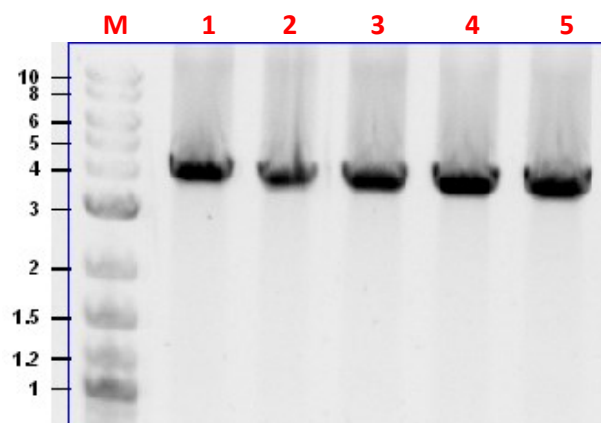


Figure S45. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvC$ pSW3_ *lacI*⁺ pACG_XylSPm_ *ilvC* clones by using primers InFusion1_seq_F17 an InFusion1_seq_R4. The expected PCR product size for positive clones is 3909 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvC$ pSW3_ *lacI*⁺ pACG_XylSPm_ *ilvC* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

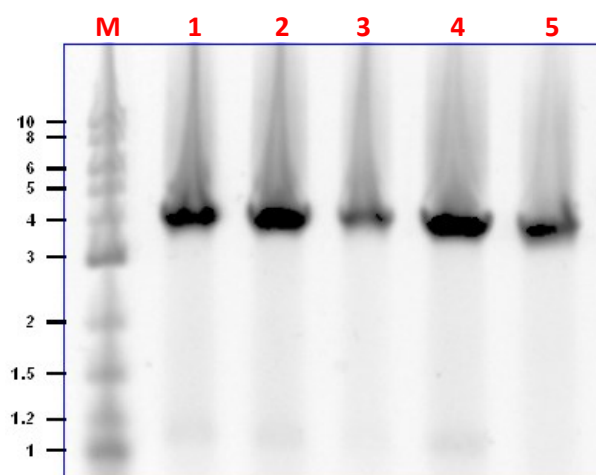


Figure S46. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta leuA$ pSW3_ *lacI*⁺ pACG_XylSPm_ *leuA* clones by using primers InFusion1_seq_F17 an InFusion1_seq_R4. The expected PCR product size for positive clones is 4005 bp. 1-5: *E. coli* K-12 BW25113 $\Delta leuA$ pSW3_ *lacI*⁺ pACG_XylSPm_ *leuA* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

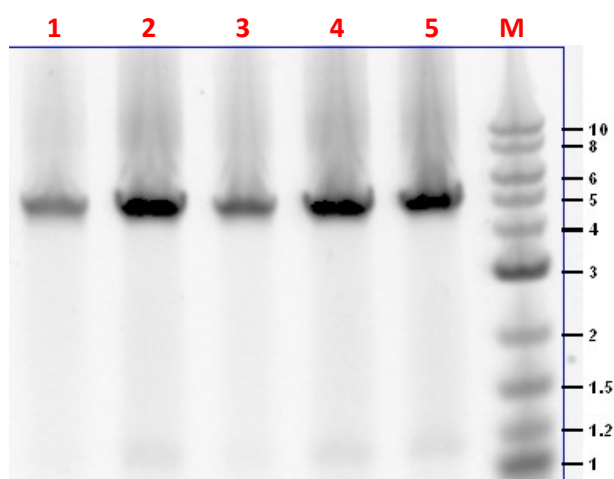


Figure S47. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta thrA$ pSW3_ *lacI*⁺ pACG_XylSPm_ *thrA* clones by using primers InFusion1_seq_F17 an InFusion1_seq_R4. The expected PCR product size for positive clones is 4896 bp. 1-5: *E. coli* K-12 BW25113 $\Delta thrA$ pSW3_ *lacI*⁺ pACG_XylSPm_ *thrA* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

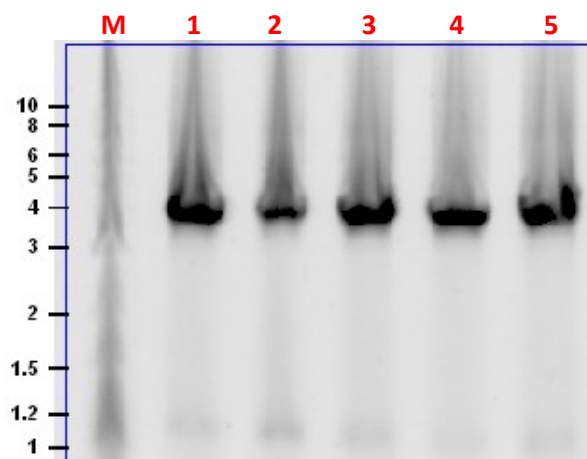


Figure S48. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvA$ pSW3_ *lacI*⁺ pACG_XylSPm_ *ilvA* clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for positive clones is 3978 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvA$ pSW3_ *lacI*⁺ pACG_XylSPm_ *ilvA* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).

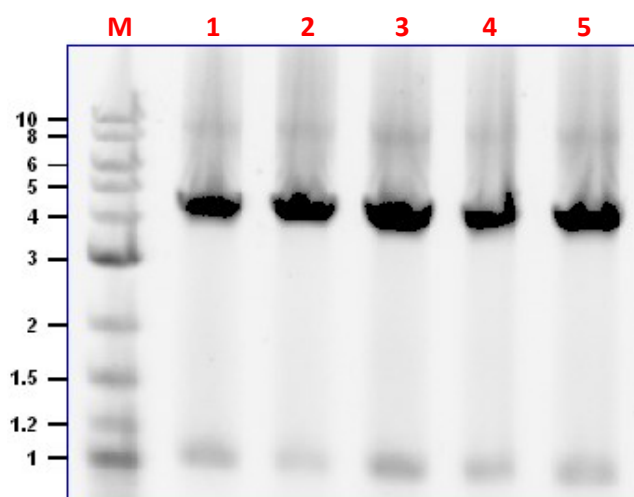
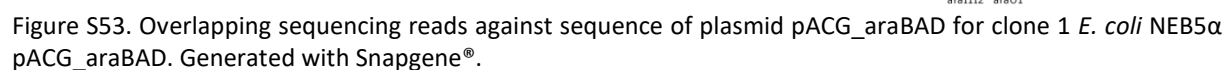
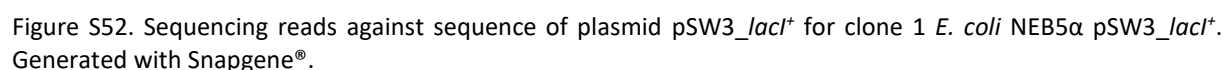
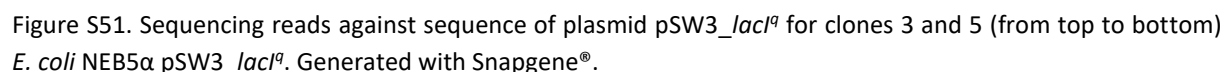
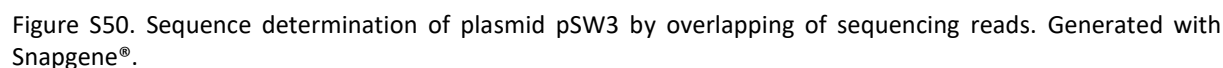


Figure S49. Gel electrophoresis of PCR products obtained by colony PCR of 5 potential *E. coli* K-12 BW25113 $\Delta ilvBN$ pSW3_ *lacI*⁺ pACG_XylSPm_ *ilvBN* clones by using primers InFusion1_seq_F17 and InFusion1_seq_R4. The expected PCR product size for positive clones is 4416 bp. 1-5: *E. coli* K-12 BW25113 $\Delta ilvBN$ pSW3_ *lacI*⁺ pACG_XylSPm_ *ilvBN* clones 1-5; M: molecular weight marker (Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)).



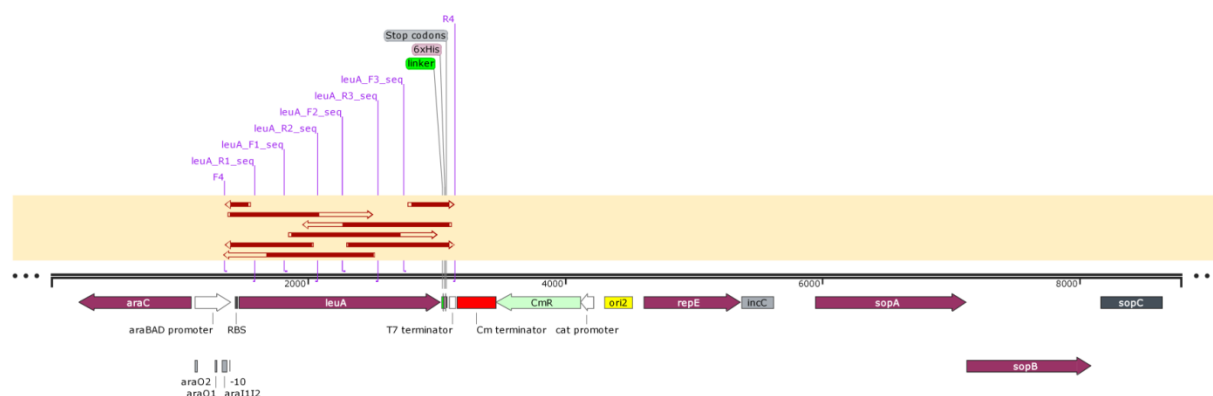


Figure S54. Overlapping sequencing reads against sequence of plasmid pACG_araBAD_levA for clone 2 *E. coli* NEB5α pACG_araBAD_levA. Generated with Snapgene®.

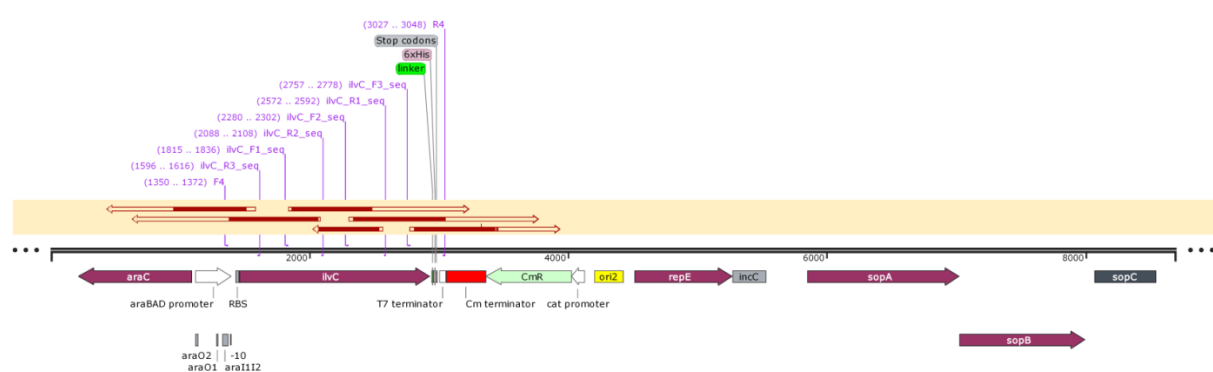


Figure S55. Overlapping sequencing reads against sequence of plasmid pACG_araBAD_ilvC for clone 4 *E. coli* NEB5α pACG_araBAD_ilvC. Generated with Snapgene®.

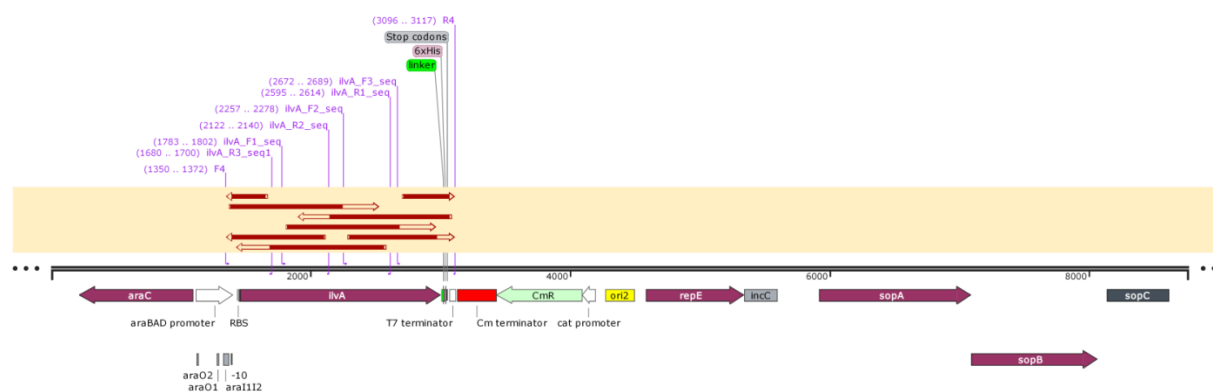


Figure S56. Overlapping sequencing reads against sequence of plasmid pACG_araBAD_ilvA for clone 5 *E. coli* NEB5α pACG_araBAD_ilvA. Generated with Snapgene®.

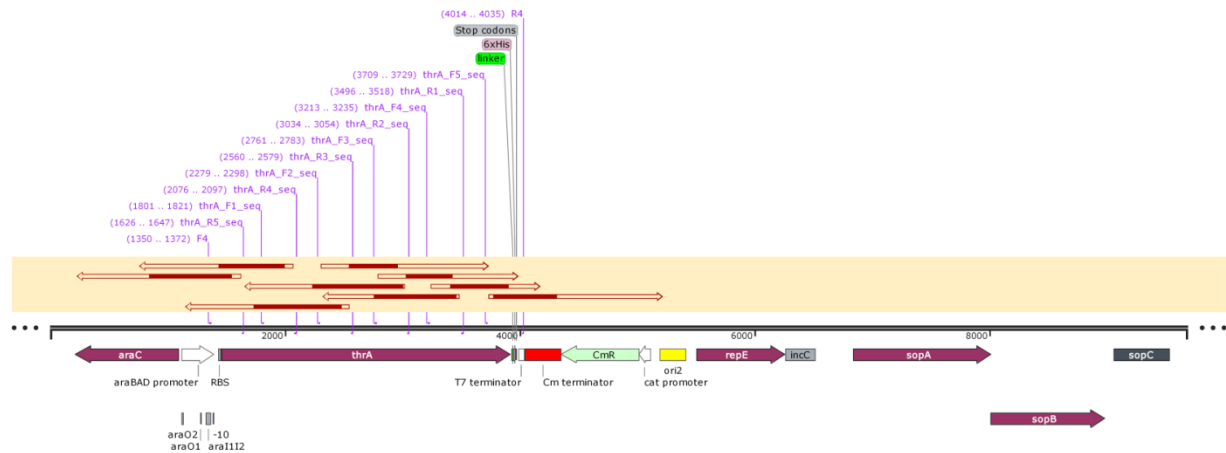


Figure S57. Overlapping sequencing reads against sequence of plasmid pACG_araBAD_thrA for clone 3 *E. coli* NEB5α pACG_araBAD_thrA. Generated with Snapgene®.

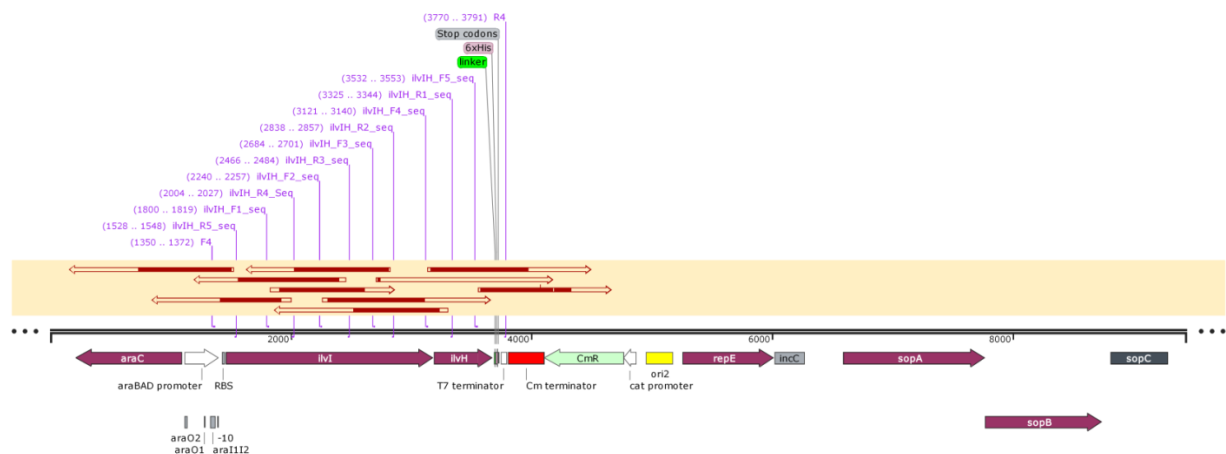


Figure S58. Overlapping sequencing reads against sequence of plasmid pACG_araBAD_ilvIH for clone 1 *E. coli* NEB5α pACG_araBAD_ilvIH. Generated with Snapgene®.

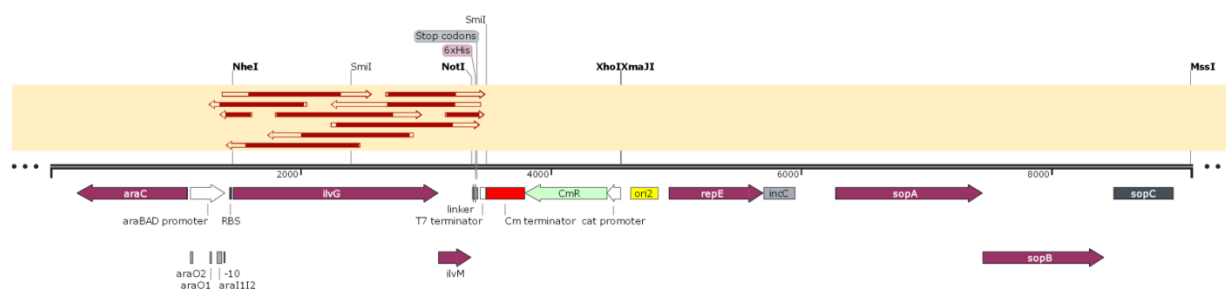


Figure S59. Overlapping sequencing reads against sequence of plasmid pACG_araBAD_ilvGM for clone 1 *E. coli* NEB5α pACG_araBAD_ilvGM. Generated with Snapgene®.

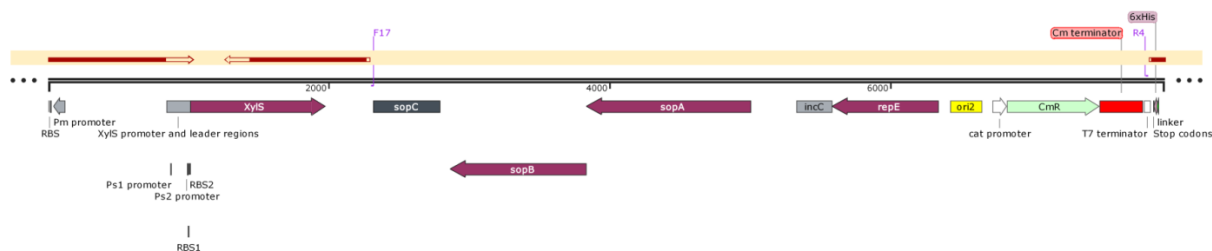


Figure S60. Sequencing reads against sequence of plasmid pACG_XylSPm for clone 16 *E. coli* NEB5α pACG_XylSPm. Generated with Snapgene®.

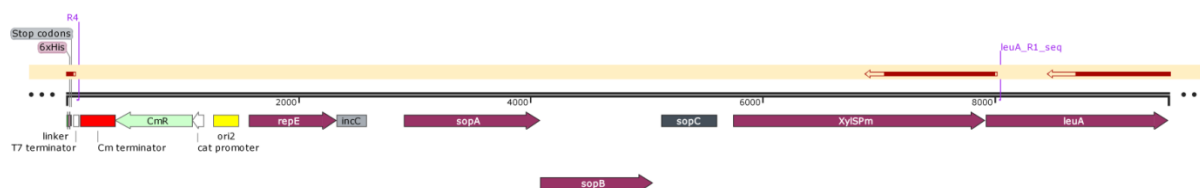


Figure S61. Sequencing reads against sequence of plasmid pACG_XylSPm_leuA for clone 4 *E. coli* NEB5α pACG_XylSPm_leuA. Generated with Snapgene®.

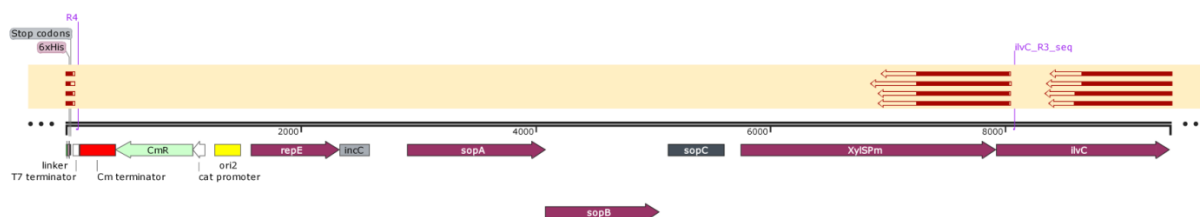


Figure S62. Sequencing reads against sequence of plasmid pACG_XylSPm_ilvC for clone 1, 3, 6 and 9 (from top to bottom) *E. coli* NEB5α pACG_XylSPm_ilvC. Generated with Snapgene®.

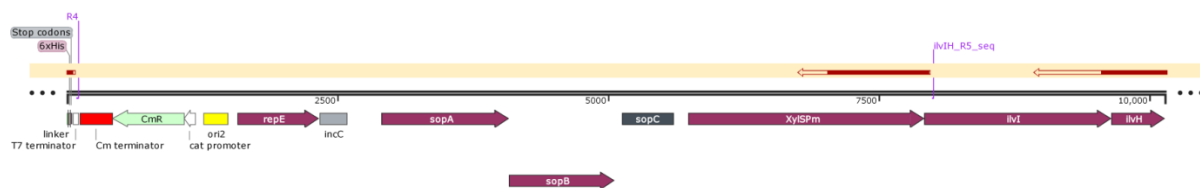


Figure S63. Sequencing reads against sequence of plasmid pACG_XylSPm_ilvIH for clone 21 *E. coli* NEB5α pACG_XylSPm_ilvIH. Generated with Snapgene®.

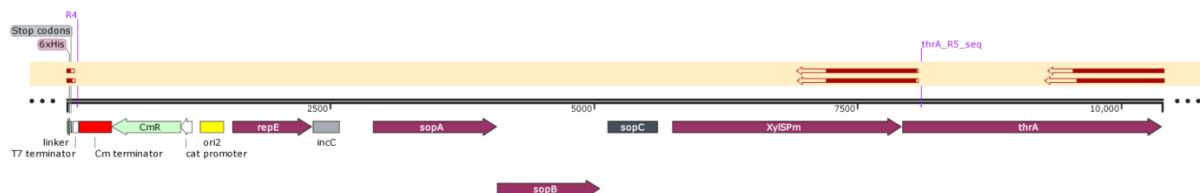


Figure S64. Sequencing reads against sequence of plasmid pACG_XylSPm_thrA for clones 2 and 8 (from top to bottom) *E. coli* NEB5α pACG_XylSPm_thrA. Generated with Snapgene®.

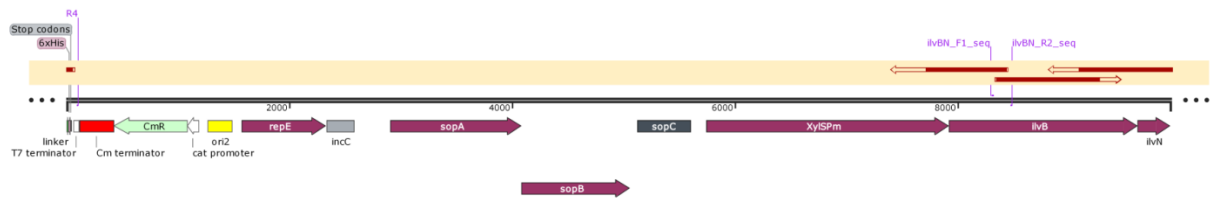


Figure S65. Sequencing reads against sequence of plasmid pACG_XylSPm_ilvBN for clone 3 *E. coli* NEB5α pACG_XylSPm_ilvBN. Generated with Snapgene®.

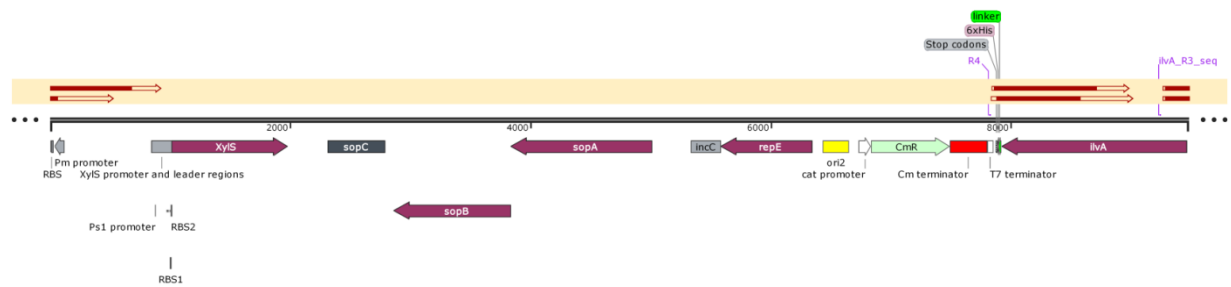


Figure S66. Sequencing reads against sequence of plasmid pACG_XylSPm_ilvA for clones 4 and 6 (from top to bottom) *E. coli* NEB5α pACG_XylSPm_ilvA. Generated with Snapgene®.

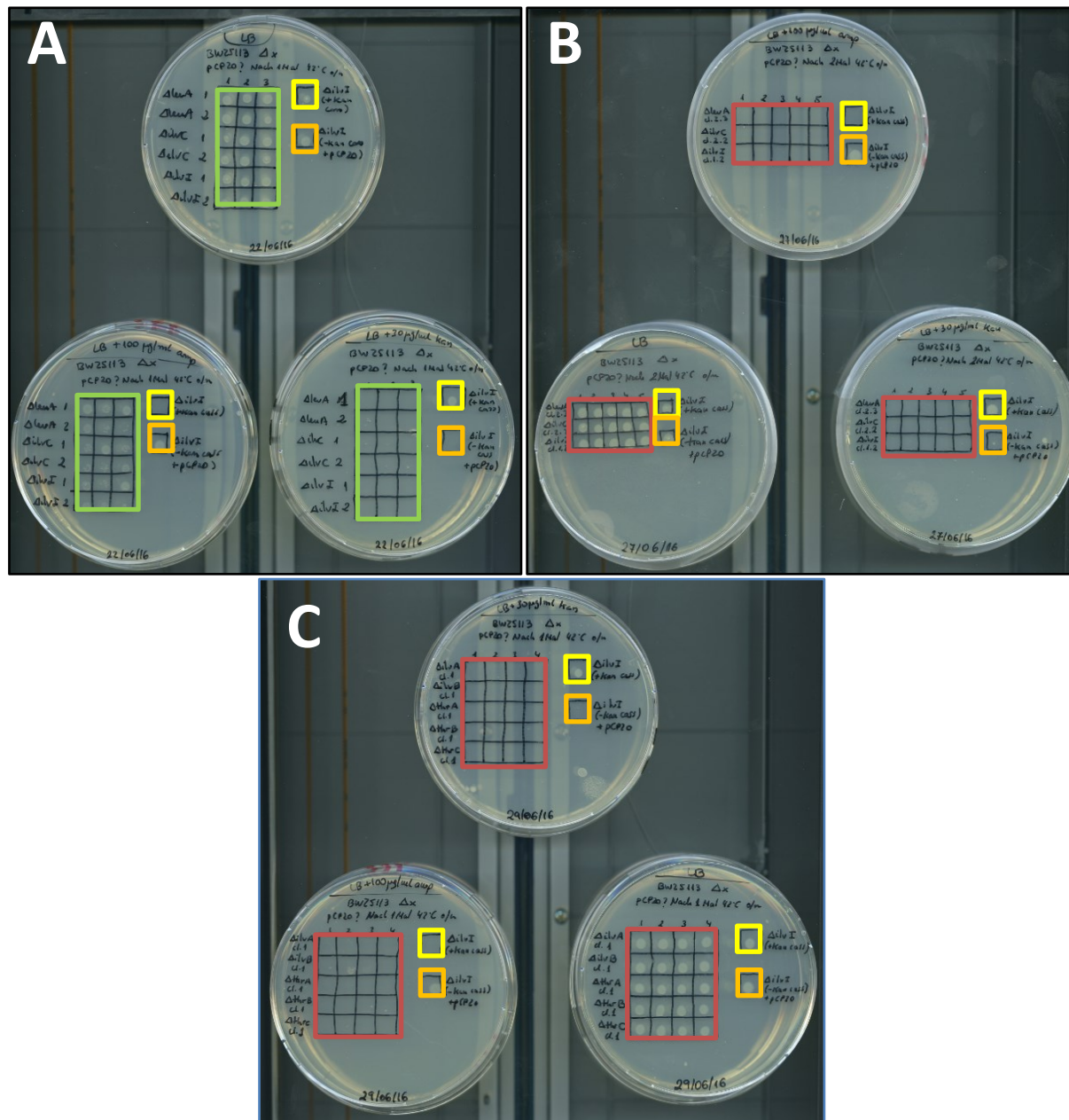


Figure S67. Verification of plasmid pCP20 curation in strains *E. coli* K-12 BW25113 $\Delta leuA$, $\Delta ilvC$ and $\Delta ilvI$ (A, B) and *E. coli* K-12 BW25113 $\Delta ilvA$, $\Delta ilvB$, $\Delta thrA$, $\Delta thrB$ and $\Delta thrC$ (C). A single 42 °C treatment was used in (A) and (C) while a second 42 °C treatment was additionally carried out in (B). For each clone, cell material was in parallel streaked out in 3 different plates: LB containing 100 µg/mL ampicillin, LB containing 30 µg/mL kanamycin and LB without antibiotics. Kanamycin sensitivity indicates that kanamycin resistance marker was successfully removed from the genome. Ampicillin sensitivity indicates loss of plasmid pCP20. Accordingly, clones growing in LB but not in antibiotic-containing plates have the correct phenotype after complete curation of pCP20 (indicated in red). Clones growing in LB but partially in LB-ampicillin (indicated in green) did not lose all copies of pCP20 and need to be temperature-treated again. As a control for selection following strains were used: *E. coli* K-12 BW25113 $ilvI::kanR$ (indicated in yellow, kanamycin resistant) and *E. coli* K-12 BW25113 $\Delta ilvI$ pCP20 (indicated in orange, ampicillin resistant).

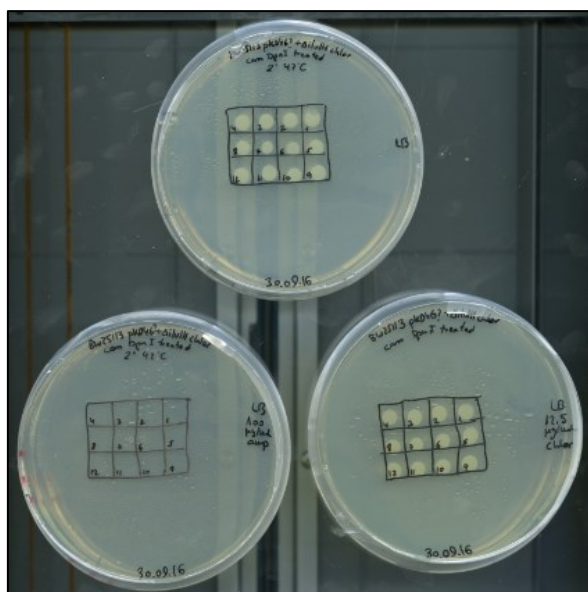


Figure S68. Verification of plasmid pKD46 curing after temperature treatment in subclones originating from clone 1 *E. coli* K-12 BW25113 *ilvH:cmR* pKD46. For each subclone, cell material was in parallel streaked out in 3 different plates: LB containing 100 µg/mL ampicillin (lower-left), LB containing 12.5 µg/mL chloramphenicol (lower-right) and LB without antibiotics (upper). Ampicillin sensitivity indicates loss of plasmid pKD46. Accordingly, clones growing in LB and LB-chloramphenicol but not in LB-ampicillin have the correct phenotype after complete curing of pKD46.



Figure S69. Verification of plasmid pCP20 curing in strain *E. coli* K-12 BW25113 *ΔilvH*. For each clone, cell material was in parallel streaked out in 3 different plates: LB containing 100 µg/mL ampicillin, LB containing 12.5 µg/mL chloramphenicol and LB without antibiotics. Chloramphenicol sensitivity indicates that chloramphenicol resistance marker was successfully removed from the genome. Ampicillin sensitivity indicates loss of plasmid pCP20. Accordingly, clones growing in LB but not in antibiotic-containing plates have the correct phenotype after complete curing of pCP20.

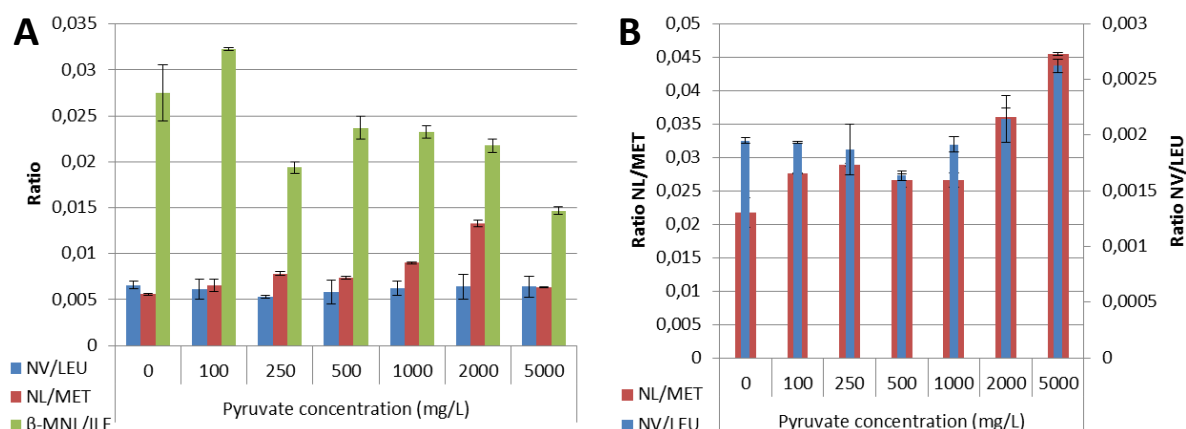


Figure S70. Calculated molar ratios of the ncBCAA with respect to the canonical counterparts in the intracellular soluble protein fraction (A) and inclusion body fraction (B) from samples taken 3.5 h after IPTG induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivations in shake flasks under cultivation conditions subjected to pulses of different pyruvate concentration. When cultivation achieved OD_{600nm} of 0.3, pyruvate was supplemented in a pulse-based manner each 30 minutes for a time period of 2.5h, this is a total of 6 pyruvate pulses. Depending on the cultivation pyruvate pulses were added up to 0, 100, 250, 500, 1000, 2000 and 5000 mg/L. Induction with 0.5 mM IPTG was performed when cultures reached OD_{600nm} of 0.6.

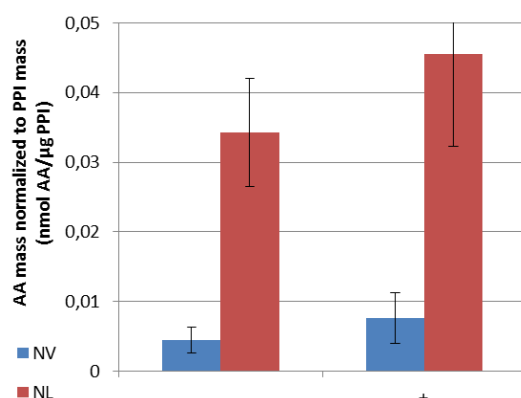


Figure S71. Molar concentrations of norvaline (blue bars) and norleucine (red bars) normalized to PPI mass present in the inclusion body fraction at 3h after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation in a 10 mL mini-reactor under standard cultivation conditions (-) or under conditions triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+).

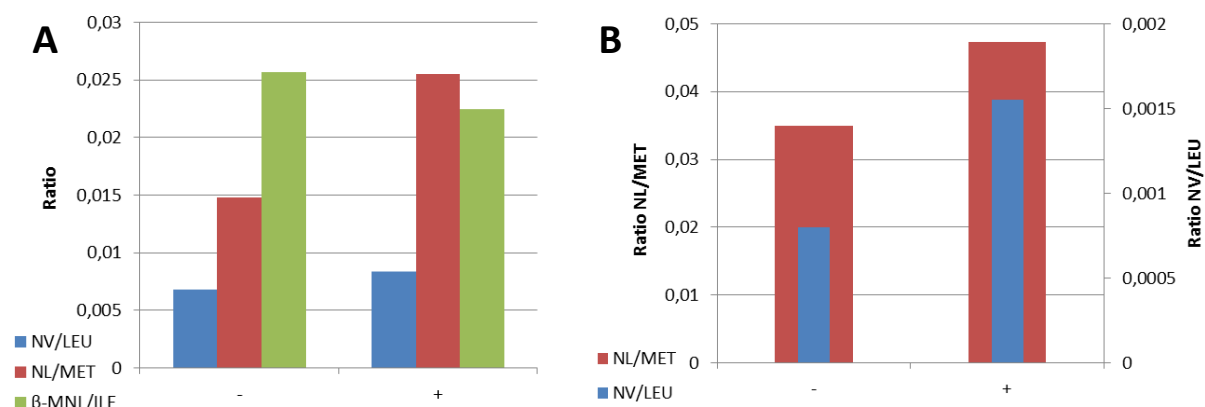


Figure S72. Calculated molar ratios of the ncBCAA with respect to the canonical counterparts in the intracellular soluble protein fraction (A) and inclusion body fraction (B) at 3h after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation in a 10mL mini-reactor under standard cultivation conditions (-) or under conditions triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+).

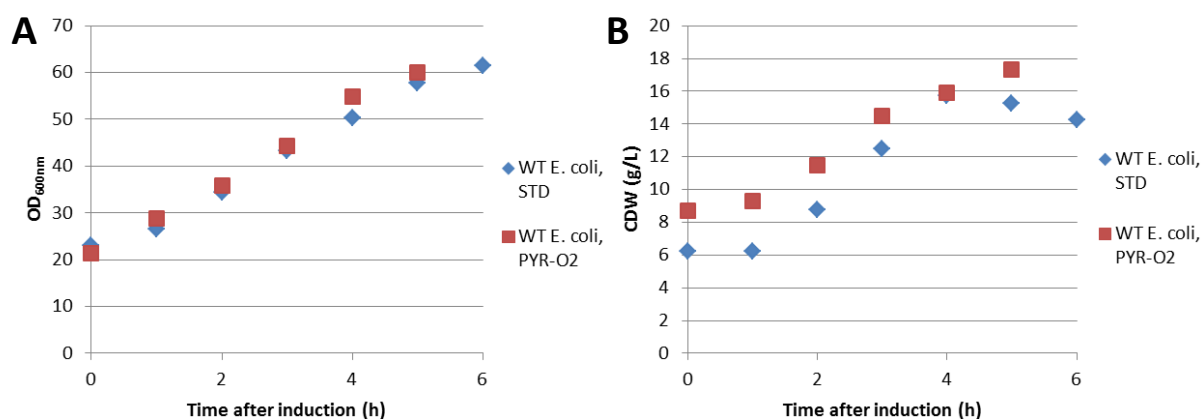


Figure S73. OD_{600nm} (A) and CDW (B) over time after induction of *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ cultivation (WT *E. coli*) in a 15L reactor under standard cultivation conditions (STD) or under conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O₂).

Table S8. Estimated recombinant mini-proinsulin concentrations ($\mu\text{g PPI}/\text{OD}0.25$) for each strain under the different tested cultivation conditions in the 10mL mini-reactor. This data correspond to the mini-reactor plate shown in Table 26. Two values are shown for each strain and condition: the first refers to standard cultivation conditions while the second applies to cultivation conditions triggering ncBCAA formation, i.e. pyruvate pulsing and O_2 limitation. Depending on the sample, analysis was done at different time points after induction: 2h AI (2h after induction) or 3.5h AI (3.5h after induction).

Strain	L-arabinose concentration (%)					
	0	0.1	0.2	0.4	0.8	1.6
<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺	0.92/0.91 (3.5h AI)	-	-	-	-	-
	0.82/0.78	-	-	-	-	-
	(2h AI)	-	-	-	-	-
<i>leuA</i> -tunable <i>E. coli</i>	-	0.95/0.91 (3.5h AI)	0.97/0.99 (3.5h AI)	1.02/1.07 (3.5h AI)	-	-
<i>ilvC</i> -tunable <i>E. coli</i>	-	-	-	0.80/0.73 (2h AI)	0.80/0.82 (2h AI)	0.75/0.79 (2h AI)
<i>thrA</i> -tunable <i>E. coli</i>	-	-	-	0.76/0.84 (3.5h AI)	0.70/0.80 (3.5h AI)	0.80/0.76 (3.5h AI)

Table S9. Estimated recombinant mini-proinsulin concentrations ($\mu\text{g PPI}/\text{OD}0.25$) 2h after induction for each strain under the different tested cultivation conditions in the 10mL mini-reactor. This data correspond to the mini-reactor plate shown in Table 27. Two values are shown for each strain and condition: the first refers to standard cultivation conditions while the second applies to cultivation conditions triggering ncBCAA formation, i.e. pyruvate pulsing and O_2 limitation.

Strain	L-arabinose concentration (%)			
	0	0.05	0.2	0.8
<i>E. coli</i> K-12 BW25113 pSW3_ <i>lacI</i> ⁺	0.79/0.79	-	-	-
<i>ilvIH</i> -tunable <i>E. coli</i>	-	0.82/0.64	-	0.79/0.82
<i>ilvA</i> -tunable <i>E. coli</i>	-	0.91/0.89	0.95/0.97	0.78/0.92
<i>ilvBN</i> -tunable <i>E. coli</i>	-	0.66/0.76	0.78/0.68	0.68/0.60
<i>ilvGM</i> -tunable <i>E. coli</i>	-	0.66/-	0.65/0.71	0.59/0.77

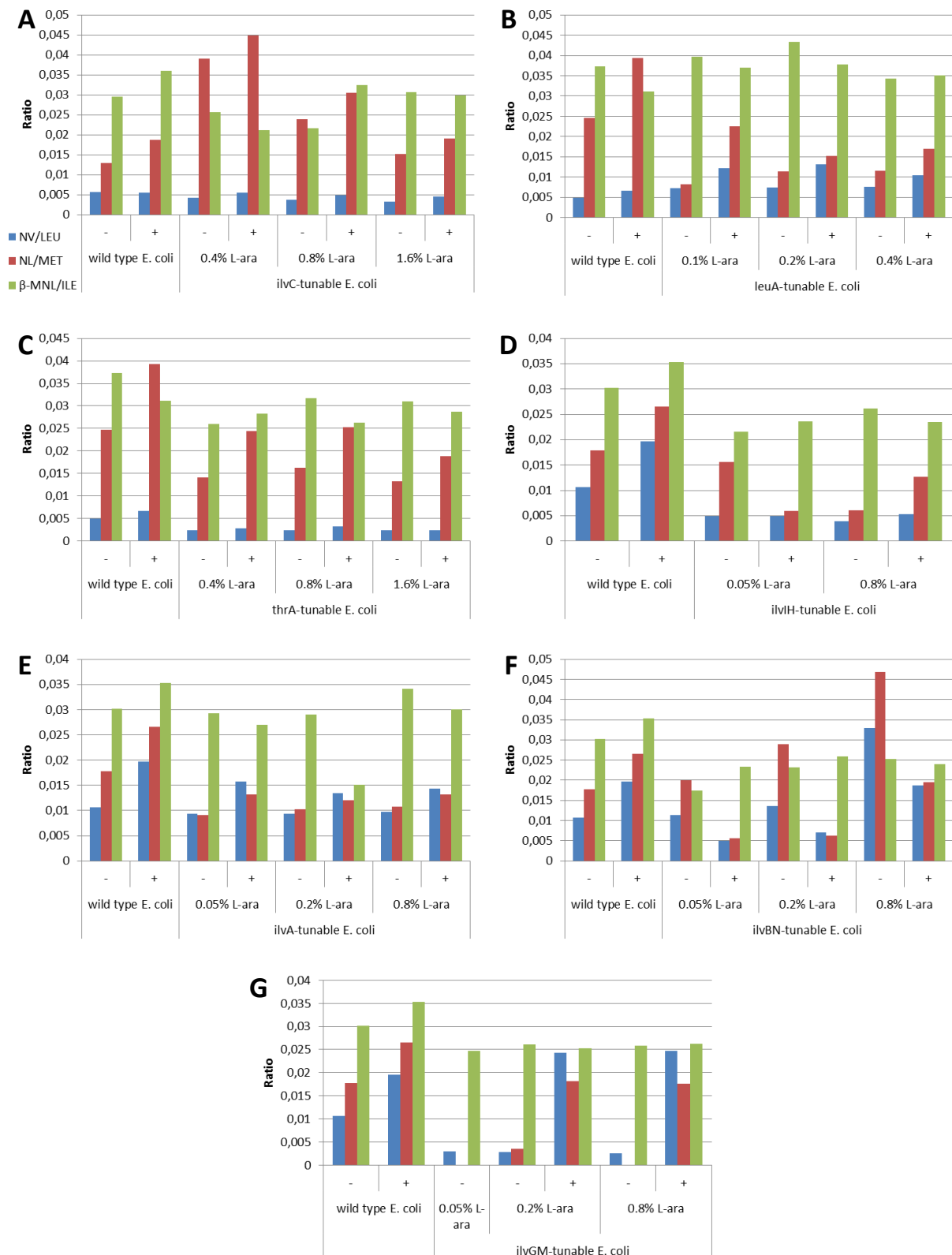


Figure S74. Calculated molar ratios of the ncBCAA with respect to the canonical counterparts in the intracellular soluble protein fraction from samples taken 2-3.5 h after IPTG induction of cultivations of *ilvC*-tunable *E. coli* (A), *leuA*-tunable *E. coli* (B), *thrA*-tunable *E. coli* (C), *ilvIH*-tunable *E. coli* (D), *ilvA*-tunable *E. coli* (E), *ilvBN*-tunable *E. coli* (F) and *ilvGM*-tunable *E. coli* (G) in a 10mL mini-reactor subjected to different L-arabinose concentrations and cultivation modes. Two cultivation modes were tested: standard cultivation conditions (-) and conditions triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+). Strain *E. coli*

K-12 BW25113 pSW3_ *lacI*⁺ (indicated as “wild type *E. coli*” in the chart) was employed as a control for comparison.

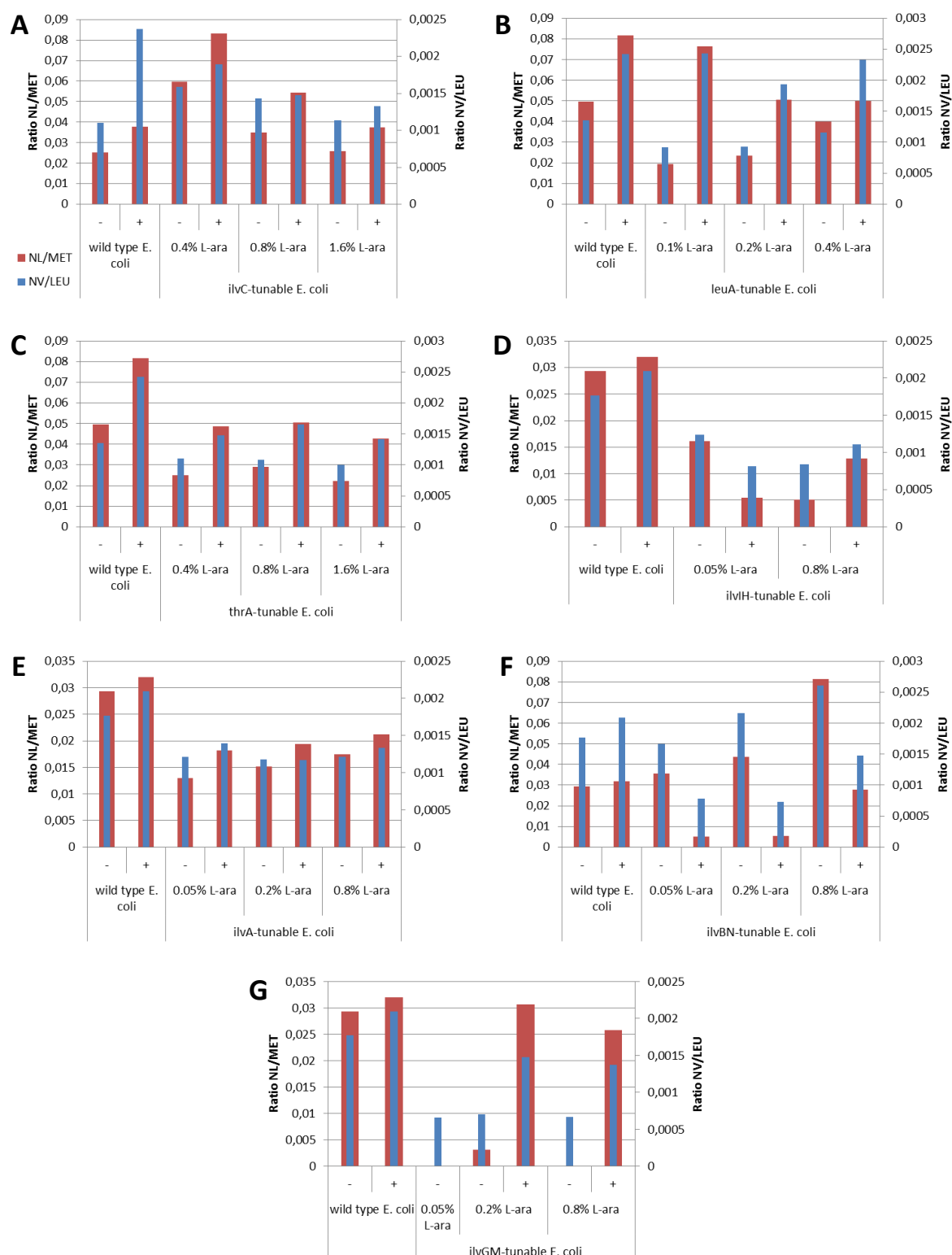


Figure S75. Calculated molar ratios of the ncBCAA with respect to the canonical counterparts in the inclusion body fraction from samples taken 2-3.5 h after IPTG induction from cultivations of *ilvC*-tunable *E. coli* (A), *leuA*-tunable *E. coli* (B), *thrA*-tunable *E. coli* (C), *ilvIH*-tunable *E. coli* (D), *ilvA*-tunable *E. coli* (E), *ilvBN*-tunable *E. coli* (F) and *ilvGM*-tunable *E. coli* (G) in a 10mL mini-reactor subjected to different L-arabinose concentrations and cultivation modes. Two cultivation modes were tested: standard cultivation conditions (-) and conditions

triggering ncBCAA production, i.e. pyruvate pulses combined with O₂ limitation (+). Strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ (indicated as wild type *E. coli* in the chart) was employed as a control for comparison.

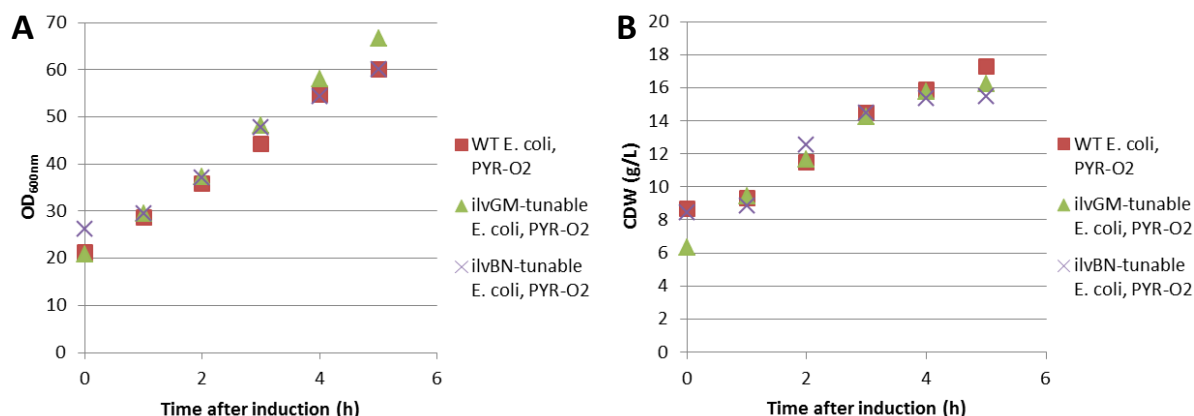


Figure S76. OD_{600nm} (A) and CDW (B) over time after induction of different *E. coli* cultivations in a 15L reactor under cultivation conditions triggering ncBCAA accumulation, i.e. pyruvate pulsing and oxygen limitation (PYR-O₂). Indicated in the legend, “WT *E. coli*” refers to the wild type strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺, “*ilvGM*-tunable *E. coli*” alludes to strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* and “*ilvIH*-tunable *E. coli*” corresponds with strain *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH*.

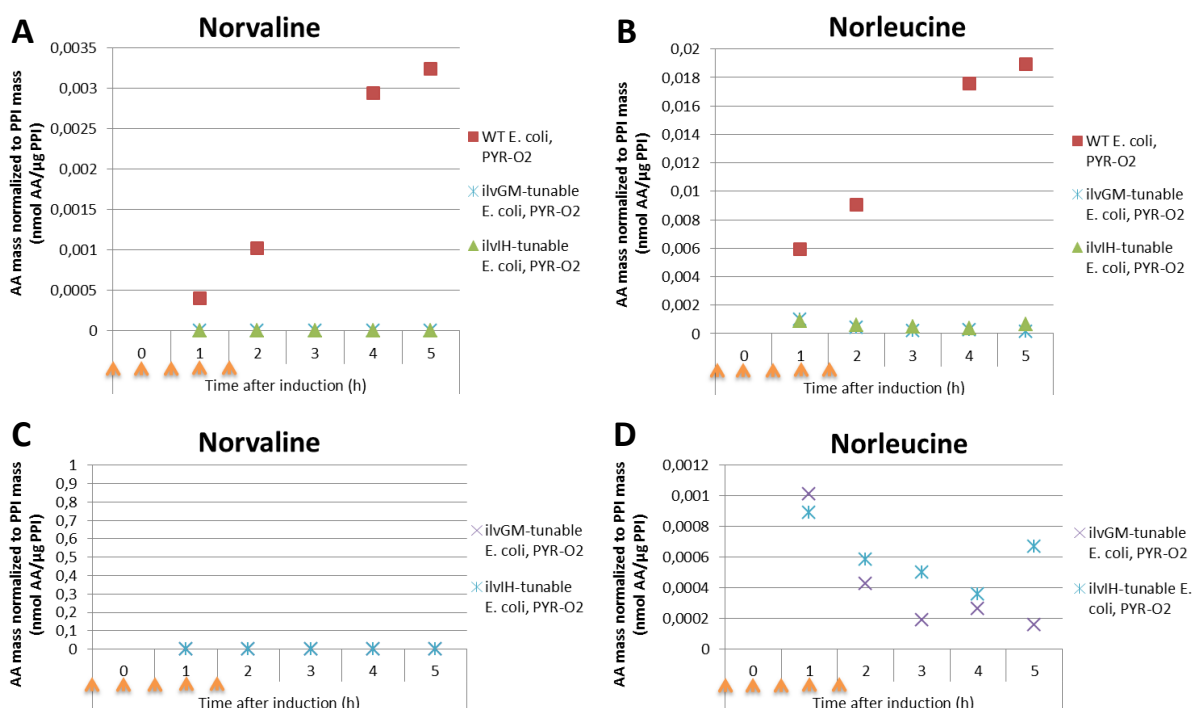


Figure S77. Molar concentrations of norvaline (A, C) and norleucine (B, D) normalized to PPI mass present in the inclusion body fraction calculated over time after induction of different *E. coli* cultivations in a 15L reactor under conditions triggering ncBCAA accumulation; i.e. pyruvate pulsing and oxygen limitation (PYR-O₂), in a big (A, B) and reduced y-axis scale (C, D). Indicated in the legend, “WT *E. coli*” refers to the wild type strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ “*ilvGM*-tunable *E. coli*” alludes to strain *E. coli* K-12 BW25113 pSW3_ *lacI*⁺ pACG_araBAD_ *ilvGM* and “*ilvIH*-tunable *E. coli*” corresponds with strain *E. coli* K-12 BW25113 Δ *ilvIH* pSW3_ *lacI*⁺ pACG_araBAD_ *ilvIH*. Orange arrows indicate time points where 1 g/L pyruvate pulse combined with 5 min O₂ limitation was applied.

ggcatgcatttacgttgacaccatcgaatggcgcaaaacctttcgcggtatggcatgatagcggcgaagagagtcaattcagggtggtgaatg
tgaaaccagtaacgttatacgaatgtcgcagagtatgcggtgtcttctatcagaccgtttcccgctggtgaaccaggccagccacgtttctgcga
aaacgcgggaaaaagtggagcggcgatggcggagctgaattacattcccaaccgctggcacaacaactggcgggcaaacagtcgttgcga
ttggcgttgccacctcagctggccctgcacgcgctcgcaaatgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtgg
tgtcgatggtagaacgaagcggcgctgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtggtgatcattaactatccg
ctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttcggcggtatttcttgatgtctctgaccagacacccatcaacagtattat
tttctccatgaagacggtacgcgactggcgctggagcatctggtcgattgggtcaccagcaaactcgcgctgttagcgggccattaaagtctgt
ctcggcgctctcgtctggctggctggcataaatactcactcgcaatcaaattcagccgatagcggaaacgggaaggcgactggagtccatgt
ccggttttaacaaaccatgcaaatgctgaatgagggcatcttccactgcgatgctggttccaacgatcagatggcgctgggcgcaatgcgc
gccattaccgagtcgggctgcggttggtcggtatctcggtagtgggatacgcgataccgaagacagctcatgttatatcccgccgttaacc
accatcaaacaggattttcgctgtctggggcaaaccagcgtggaccgcttgcgcaactctctcagggccaggcgtgaagggaatcagctgtt
gcccgtctactggtgaaaagaaaaaccaccctggtgcccaatacgcgaacccgctctccccgcgcttgccgattcattaatgcagctggcac
gacaggtttcccgactggaaagcgggcagtgagcgcaacgcgcgattgaggaccgcatctgcgctgggcttggtctcgcgagcatcagcgcgt
cagtgccgtgcatcacgacaccgacaacctgcacatccatctcgccatcaacaagattcaccgacccgaaacaccatccatgagccgtatcgg
gcctaccgcgcccctgcgtgacctctgcgcgacgctgaacgggactacgggcttgagcgtgacaatcacgaaacgcggcagcgc

Figure S78. Nucleotide sequence of the genetic region of *E. coli* K-12 BW25113 genome containing the gene *lacI* and its corresponding *lacI*⁺ promoter variant. Blue: promoter sequence corresponding to the *lacI* promoter. Red: single base offering *lacI*⁺ phenotype. Green: coding region of gene *lacI*.

Recent dissertations at Bioprocess Engineering, TU Berlin

1. Sarah Kamel Hassan (2019) Enzymatic synthesis of α -D-pentofuranose-1-phosphate as a universal glycosyl donor for modified nucleosides synthesis using thermostable nucleoside phosphorylase
2. Funda C. Erterm (2019) Life cycle assessment and modelling approaches as a combined evaluation tool for sustainable control strategies at biogas plants
3. Emmanuel Anane (2019) Model-based strategies for scale-down studies in fed-batch cultivation of *Escherichia coli* expressing proinsulin
4. Martin Kögler (2018) Advanced Raman spectroscopy for bioprocess monitoring
5. Anna Maria Marbá (2018) Monitoring of the single-cell morphology for the evaluation of microbial eukaryotic bioprocesses
6. Juan Antonio Arzate (2018) Modelling and simulation of biogas production based on anaerobic digestion of energy crops and manure
7. Anika Bockisch (2018) Mobile multi-parameter measurements for the dynamic analysis of gradients in brewing vessels
8. Elvis Legala (2018) Reconstruction of the lanthipeptide ruminococcin-A biosynthesis machinery in *Escherichia coli* and structural characterization
9. Ekaterina Osmekhina (2018) Quantitative approach for detection of RNA and proteins by sandwich hybridization technology
10. Anja Lemoine (2017) Impact of oscillations in substrate and oxygen availability on *Corynebacterium glutamicum* cultivations
11. Basant El Kady (2017) Application of the enzyme controlled glucose feeding for the production of recombinant proteins. Starter cultures and autoinduction
12. Julia Glazyrina (2017) Endophytes as source of polypeptide and polyketide antibiotics fengycin and bacillaene. Process development for fengycin production
13. Florian Glauche (2017) Platform technologies for automated bioprocess development
14. Eva Brand (2017) Growth rate dependent impact of oscillating conditions on *Escherichia coli* physiology