
University of Lille, Sciences and Technologies

Doctoral School of Material Science, Radiation and Environment (ED SMRE)

PhD Thesis

Order No.: 42292

for the degree of

Doctor of Philosophy (PhD)

(Specialization: Biological Function Engineering)

by

Debarun DHALI

Correlation between lipopeptide biosynthesis and their precursor metabolism in *Bacillus subtilis*

In the Charles Viollette Institute, EA 7394

Defended on 15 December, 2016

CONFIDENTIAL

Committee members:

Valérie LECLERE, Assistant Professor	University of Lille, Sciences and Technologies, Lille, France	Examiner
Philippe JACQUES, Professor	University of Lille, Sciences and Technologies, Lille, France	Director
Manuel CANOVAS, Professor	University of Murcia, Spain	Reviewer/President
Michael LALK, Professor	University of Greifswald, Germany	Reviewer
François COUTTE, Lecturer	University of Lille, Sciences and Technologies, Lille, France	Examiner
Matthieu JULES, Researcher Associate	INRA, Jouy en Josas, France	Examiner

Université de Lille, Sciences et Technologies
Ecole Doctorale: Science de la Matière du Rayonnement et de l'Environnement (ED SMRE)

Thèse

Ordre No.: 42292

Pour l'obtention du grade de

Docteur en Philosophie (PhD)

(Spécialité: Ingénierie des Fonctions Biologiques)

par

Debarun DHALI

Corrélation entre la biosynthèse des lipopeptides et le métabolisme de leurs précurseurs chez *Bacillus subtilis*

Préparée à l'Institut Charles Viollette, EA 7394

Défendue le 15 Décembre, 2016

CONFIDENTIEL

Jury

Valérie LECLERE, Professeur Assistant	Université de Lille, Sciences et Technologies, Lille, France	Examinatrice
Philippe JACQUES, Professeur	Université de Lille, Sciences et Technologies, Lille, France	Directeur
Manuel CANOVAS, Professeur	Université de Murcia, Espagne	Rapporteur/Président
Michael LALK, Professeur	Université de Greifswald, Allemagne	Rapporteur
François COUTTE, Maître de Conférences	Université de Lille, Sciences et Technologies, Lille, France	Examineur
Matthieu JULES, Researcher Associate	INRA, Jouy en Josas, France	Examineur

*To Ma and Papa
My pillar of support*

“Be less curious about people
And more curious about ideas”
- Marie Curie

“Take risks in your life,
If you win, you can lead
If you lose, you can guide”
- Swami Vivekananda

ACKNOWLEDGMENTS

The journey towards a PhD is by itself an adventure. You keep falling but always to strike back stronger. This strength comes from the wonderful people that were in many ways an integral part of my journey. Of course, this started with the guidance, support and constant encouragement of my supervisor and mentor, the director of my thesis, **Prof. Philippe Jacques**. His motivations and insightful guidance are one of the main reasons leading to the successful completion of the work. I would like to express my heartfelt gratitude for believing in me, which was one of the main factors, which always kept me going.

My other driving force was **Dr. François Coutte**, my co-supervisor, whom I am immensely grateful to for his unconditional support throughout the three years. He was always available for insightful discussions whenever I needed his presence. His dynamic attitude towards work has kept inspiring me even during my failures and motivated me to thrive further. I would like to thank him for helping me right from the day I arrived in this city.

I am thankful to **Dr. Joachim Niehren** and **Dr. Cristian Versari** for the development of the metabolic network and providing interesting prediction which really helped me a lot.

I am thankful to **Prof. Pascal Dhulster** for allowing me to access the laboratory and research facilities.

My work and stay at INRA was meaningful due to the assistance of **Dr. Sandrine Auger** and **Dr. Vladimir Bidenko**. I am really grateful to **Dr. Vladimir Bidenko** for providing lots of in-depth knowledge of Pop in – Pop out technique as well as helping in getting my first meaningful achievement. I am really thankful to him as he drove me to the lab even on Saturdays and Sundays so I can work and get the results required. I am thankful to the members in Gembloux for LC-MS analysis.

I am thankful to **Prof. Michael Lalk** for allowing me to work in his laboratory and accepting to review my thesis. I am also thankful to **Prof. Manuel Canovas** for accepting to review my thesis. And, I would like to express gratitude to **Dr. Valerie Leclere** and **Dr. Matthieu Jules** for accepting to be the jury members for my defense.

I would also like to thank the entire staff at ProBioGEM, for providing me with a friendly environment to work in. I cherish the valuable time they spent with me. Especially I would like to thank **Dr. Max Béchet** for helping me in the molecular biology experiment in the early days. I am really grateful to **Corine** for helping me every time I needed her help for

the experiments. I am thankful to **Cathy** for clearing my bills every time I travelled. I really appreciated all of their help. I have learnt from them that if you have the eagerness to help, language cannot be a barrier.

Besides the staff, I am thankful to **Juliette, Remi, Antoine, Luiz, Qassim, Oumayma, Mickael, Kalim, Sandy, Delphine, Alaa, Ameen** and all other members in C124 for their time, each time I needed them. Alongside my lab members in Lille, I am grateful to all the fellows (**Laura, Transi, Jad, Nora, Paola, Joana, Nicolas, Luminita** and **Maria del Carmen**) in AMBER project for their discussion on facebook and whenever we met. **Joana** really helped me during my stay in Greifswald and helped me a lot to perform and learn the technique to analyze extracellular metabolites. **Paola** also helped with my work during my stay in INRA. I am thankful to **Hannah** for assisting me in the last phase of my experiments.

This journey wouldn't have been the same without the help, support and company of my dearest friends **Narinder, Mayank, Mohit, Samadhan, Pushpendra** and **Neetu** for being amazing friends and keeping company with me in everything. I really enjoyed the friendly chats with my coffee partners **Yazeen** and **Thibault** as well as for their help in experiments and many academic related matters. Enjoyment in Lille would not have been the same without the party couple **Koyal** and **François**; I really had some memorable moments with them.

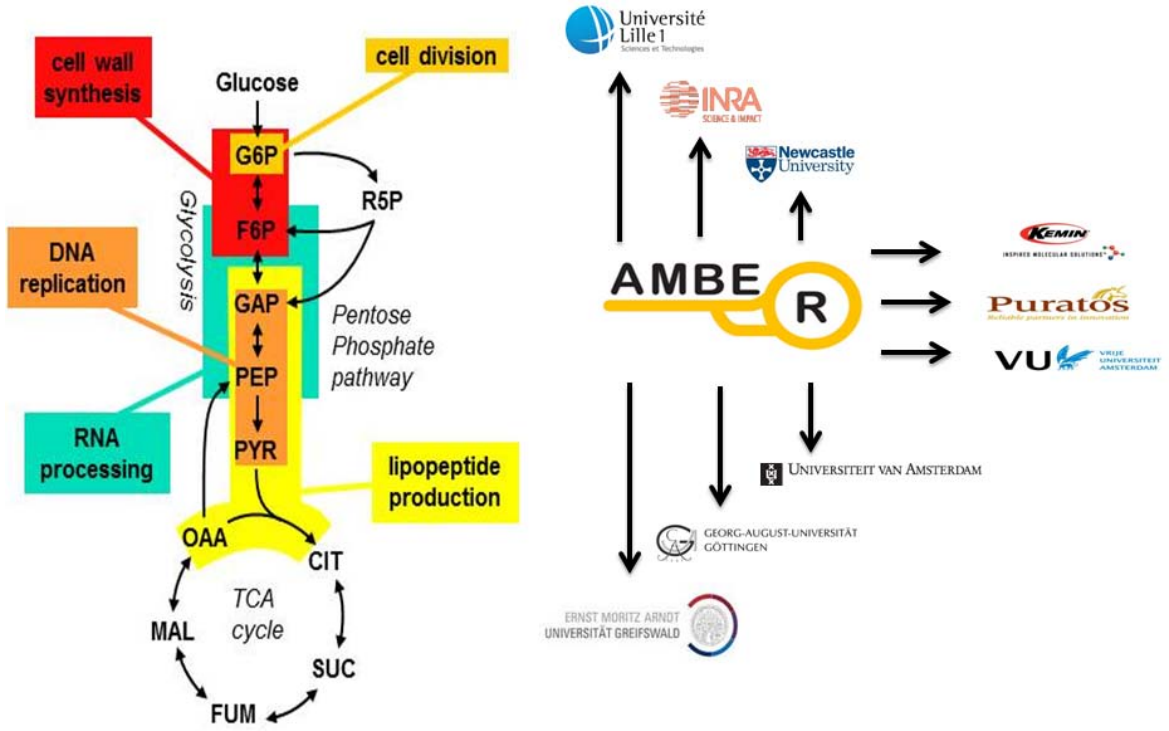
I would like to express my deepest love and gratitude to **Ma** and **Papa** for being there and bearing me in all times. Their shower of love and affection helped to maintain positivity in my life during this period. You are the reason for whatever little things I have achieved in life. I dedicate my thesis to the both of you.


Shreya has been the closest person to me in my life besides my parents. She has been a source of encouragement, knowledge and intellect. I am really thankful to have her as my life partner. She helped me a lot during this entire journey. I am really grateful to her for the constant support and especially bearing with my temper all the time.

I am thankful to **European Union** and **Marie Curie ITN AMBER (317338)** for funding this project.


Lastly, I am thankful to **Lille**, it is my home away from home.


AMBER Network






EU FP7 Marie Curie Initial Training Network on Molecular Bacteriology







ABSTRACT

Genetic and metabolic engineering approaches were implemented to increase the precursor availability required for the biosynthesis of surfactin, a lipopeptide produced by *Bacillus subtilis*. This study is based on the overproduction of surfactin through enhancement of intracellular pool of leucine. A metabolic network of branched chain amino acids (BCAA: isoleucine, leucine and valine) metabolism in *B. subtilis* was first formalized through biocomputing tools. Predictions were based on genes which are directly or indirectly involved in the regulation of the promoter ($P_{ilv-leu}$) activity of the *ilv-leu* operon in the final step of leucine biosynthesis or influx of pyruvate (glycolytic pathway). A markerless gene deletion strategy (PopIn- PopOut technique) was adopted to carry out multiple deletions in a single strain and avoid antibiotic limitation. Various single knockout strains were developed with the deletion of *bcd*, *codY*, *lpdV* and *tnrA* genes. Besides these single knockout genes, a double knockout strain with the deletion of *codY* and *tnrA* was developed too. Strains with the deletion of the binding region of CodY (site II) and a leucine T-box regulator were also studied. It was observed that the deletion of these genes have varied positive impact on surfactin production, quantitatively and qualitatively. The role of extracellular metabolites in determining the production of surfactin isomers was analyzed. Genes (*pgcA*, *pgi*, *pfkA* and *pykA*) which encode various enzymes in the glycolytic pathway were also deleted, to study the role of pyruvate metabolic flux in the surfactin production. Gene *yvbW*, a potential candidate for leucine with a hypothetical function as leucine transporter was also studied for the effect on surfactin production. Moreover, the surfactin production of all these mutants was analyzed in different environmental conditions (medium composition and aeration). Depending on the environmental conditions and the gene knockout, the surfactin production was enhanced with a factor from 2 to 21. This work establishes that the precursor limitation problem of lipopeptide biosynthesis can be overcome by using this integrative approach.

RÉSUMÉ

Dans ces travaux des approches de génie génétique et métabolique ont été mises en œuvre pour augmenter la fourniture des précurseurs nécessaire à la biosynthèse d'un lipopeptide produit par *Bacillus subtilis*. Cette étude a porté plus particulièrement sur la surproduction de la surfactine grâce à l'augmentation du pool intracellulaire de leucine. Le réseau métabolique du métabolisme des acides aminés branchés (BCAA: isoleucine, leucine et valine) chez *B. subtilis* a d'abord été formalisé et modélisé par des outils bio-informatique. Puis, grâce à l'utilisation de la programmation par contrainte et à l'interprétation abstraite des interruptions de gènes ont été prédites dans ce réseau. Les prédictions se sont basées sur des gènes qui sont directement ou indirectement impliqués dans la régulation de l'activité du promoteur de l'opéron *ilv-leu* ($P_{ilv-leu}$), dans les voies de dégradations des BCAA, ainsi que dans le métabolisme central du carbone. Une stratégie de suppression de gènes (knockout) sans l'utilisation de marqueur antibiotique (technique du PopIn-PopOut) a été entreprise pour mener à bien ces multiples délétions dans une même souche et en évitant ainsi une limitation par les antibiotiques. Différent mutants contenant de simple knockout des gènes *bcd*, *codY*, *lpdV*, *tnrA*, ainsi que du site de fixation de CodY sur $P_{ilv-leu}$ et de son atténuateur à leucine (T-box) ont été construits. Un mutant contenant le double knockout des gènes *codY* et *tnrA* a aussi été développé. Les résultats obtenus suite à la délétion de ces gènes ont montré un impact positif sur la production du lipopeptide tant d'un point de vue quantitatif que qualitatif. Les gènes (*pgcA*, *pgi*, *pfkA* et *pykA*) qui codent pour diverses enzymes de la glycolyse ont également été supprimés afin de confirmer le rôle du flux métabolique en pyruvate dans la production de la surfactine. Par ailleurs, la délétion du gène *yvbW*, qui est responsable du transport extracellulaire de la leucine a également montré un impact significatif sur la production de la surfactine. Enfin, la production de surfactine de tous ces mutants a été analysée dans différentes conditions environnementales (composition du milieu de culture et aération). En fonction des mutants et du milieu de culture utilisés la production surfactine a été améliorée d'un facteur de 2 à 21. Ce travail a établi que la limitation de la biosynthèse de la surfactine par la fourniture en précurseur peut être surmontée en utilisant cette approche intégrée.

Table of Contents

GENERAL INTRODUCTION	1
CHAPTER 1: LITERATURE REVIEW	2
1. <i>BACILLUS SUBTILIS</i> : IN A NUTSHELL	3
1.1 <i>Bacillus subtilis</i> and its genome	3
1.2 The complex regulation of the lipopeptide synthesis	4
1.3 Industrial applications	6
2. INSIGHTS INTO NON-RIBOSOMAL PEPTIDE SYNTHESIS (NRPS), UNIQUE MACHINERY	9
2.1 Introduction	9
2.2 Primary domains	10
2.3 Secondary domains	16
2.4 Classification	19
2.5 Precursors of the NRPS mechanism and their regulation mechanisms	21
2.6 Interest of the NRPS peptides produced by <i>B. subtilis</i>	30
3. SURFACTIN: EXPLORING FROM GENE TO PRODUCT	31
3.1 Discovery	31
3.2 Structure	31
3.3 Properties	32
3.4 Biosynthesis	33
3.5 Regulators	36
3.6 Structure and activity relationship	38
3.7 Analysis	40
3.8 Application	40
4. OVERPRODUCTION OF SURFACTIN	45
4.1 Present scenario	45
4.2 Role of various factors	45
4.3 Direct approach	50
4.4 Linkage between Direct and Indirect approach	54
4.5 Indirect approach	55
CHAPTER 2: MATERIALS AND METHODS	57
MATERIALS	58
1 Strains	58
2 Plasmids	60
3 Genotype of plasmids and <i>E.coli</i> strains	63

4 Culture Media.....	65
METHODS	66
5 DNA Manipulation.....	66
6 Formalization of reaction network	68
7 Mutant construction.....	70
8 Transformation	73
9 Surfactin analysis	74
10 LC-MS-MS analysis of surfactin isomer	76
11 Extracellular metabolome	77
CHAPTER 3: RESULTS AND DISCUSSION	78
1. Is intracellular pool of leucine a limiting factor for surfactin biosynthesis?	79
2. How to enhance leucine production within the intracellular pool ?	82
2.1 Formalization of the model of BCAA metabolic pathway	82
2.2 Knockout prediction	85
2.3 Mutant constructed	88
2.4 Growth and surfactin analysis	91
2.5 Publication and Valorization of these results	99
3. FACTORS INFLUENCING SURFACTIN PRODUCTION.....	101
3.1 Role of nutrient composition.....	101
3.2 Impact of oxygen transfer	107
3.3 Other factors.....	110
3.4 Publication and Valorization of these results	116
4. ROLE OF PYRUVATE FLUX	117
4.1 Mutant constructed	117
4.2 Growth and surfactin analysis	118
4.3 Carbon metabolism and surfactin production	125
5. IMPACT OF DELETION ON SURFACTIN VARIANTS	131
5.1 Isoform analysis	131
5.2 Extracellular metabolite analysis	133
5.3 Publication and Valorization of these results	139
6. PERMEASE.....	140
6.1 Mutant constructed	140
6.2 Growth and surfactin analysis with native promoter and constitutive promoter .	141
6.3 Publication and Valorization of these results	147
CHAPTER 4: CONCLUSION AND PERSPECTIVES	148
APPENDIX I	175
VECTORS AND PLASMIDS	175

pGEM-T Easy	175
pDG1661	175
pBC16.....	176
pUC57	176
pBG124	177
pBG127	177
Synthesized sequence (P_{repU} - <i>ermC</i> - p_{srjAA}).....	178
APPENDIX II	179
MEDIA COMPOSITIONS	179
LB MEDIUM	179
COMPETENCE MEDIUM	179
LANDY MEDIUM	180
TSS SUPPLEMENTED WITH 16 AMINO ACIDS.....	182
AMBER MEDIUM.....	183
APPENDIX III	184
APPENDIX IV	187

LIST OF FIGURES

Figure 1 Distribution of essential genes in <i>B. subtilis</i> 168 genome	3
Figure 2 Overview of regulatory pathways affecting competence development as well as positive and negative regulation of <i>comK</i> and related genes are indicated by arrows and T-bars (Hamoen et al., 2003).....	6
Figure 3 Amino acid recognition and activation by ATP hydrolysis by Adenylation (A) domain (Sieber and Marahiel, 2005)	11
Figure 4 Role of <i>sfp</i> gene for the attachment of prosthetic arm (Sieber and Marahiel, 2005)	12
Figure 5 Activated amino-acyl adenylate forming a covalent bond with the free thiol of PCP domain (Sieber and Marahiel, 2005).....	13
Figure 6 Peptide bond formation between acceptor (a) substrate and donor (d) substrate is facilitated by C-domain (Sieber and Marahiel, 2005)	14
Figure 7 Role of TE domain for peptide release (Sieber and Marahiel, 2005)	15
Figure 8 Role of epimerization (E) domain during elongation (Sieber and Marahiel, 2005)	16
Figure 9 Reaction involving N-methyltransferase (Mt) domain, the yellow box indicates the modification (Marahiel, 2009).....	17
Figure 10 Reaction involving Formylation (F) domain, the yellow box indicates the modification (Marahiel, 2009).....	18
Figure 11 Reaction involving oxidation (Ox) and reduction(R) domains, the yellow box indicates the modification (Marahiel, 2009)	18
Figure 12 Assembly and mode of synthesis of linear NRPSs (Mootz et al., 2002)	19
Figure 13 Assembly and mode of synthesis of iterative NRPSs (Mootz et al., 2002)	20
Figure 14 Assembly and mode of synthesis of nonlinear NRPSs (Mootz et al., 2002)	21
Figure 15 Role of glycolytic pathway and production of various metabolites as a result of this pathway (Tojo et al., 2010).....	22
Figure 16 Complex regulation of branched chain amino acid pathway (Brinsmade et al., 2010).....	25
Figure 17 Role of various genes for the production of isoleucine, leucine and valine (Belitsky, 2015)	26
Figure 18 The role of various enzymes in the synthesis of branched chain fatty acids	27

Figure 19 Branched chain amino acid transporters A) YvbW (both import and export of leucine); B) BcaP, BrnQ and BraB (import of atleast one BCAA) ; C) AzlCD export of BCAAs	29
Figure 20 Primary structure of surfactin, where n depends on the chain length of fatty acids (Seydlová and Svobodová, 2008).....	31
Figure 21 Operon responsible for the biosynthesis of surfactin, A: denoting the domains present in each sub-unit, B: indicates the amino acid inoculated by each subunit in the final peptide moiety (Jacques,2011)	33
Figure 22 Modular and domain representation for surfactin biosynthesis (Bruner et al., 2002).....	34
Figure 23 An overview of regulatory pathways affecting surfactin operon expression (Jacques, 2011).....	37
Figure 24 Development of master strain <i>B. subtilis</i> BSB1N (Tanaka et al., 2013).....	59
Figure 25 The integration plasmid pBG400 used for the integration of <i>sfp</i> gene	60
Figure 26 A) Plasmid K7010; B) upp-phleo-cI cassette used for markerless deletion	61
Figure 27 The integration plasmid pBG133 used for the integration of P_{repU} -neo gene.....	62
Figure 28 The integration plasmid pBG401 used for the integration of P_{repU} -ermC gene..	63
Figure 29 Various identifiers in the reaction network.....	69
Figure 30 General overview of Pop-In and Pop-Out technique	72
Figure 31 A) Biolector [®] fermentation system; B) 48-well MTP; C) Probes inside each well	75
Figure 32 General Surfactin structure: It consists of four leucine molecules among seven amino seven amino acids present within the peptidic moeity, the position of the leucine molecules are encircled	79
Figure 33 Leucine feeding : Impact on specific surfactin yield with <i>B. subtilis</i> BBG111 in two different media condition with and without the presence of leucine in the culture media at 37°C in shake flask condition with 160 rpm. Results are expressed as mean value with standard deviation.....	80
Figure 34 Branched chain amino acid reaction network. The concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm.....	82
Figure 35 PCR verification for markerless deletion 1: Δbcd ; 2 : $\Delta codY$; 3 : $\Delta lpdV$; 4: $\Delta tnrA$; M: 1 kb Ladder	88
Figure 36 Surfactin production (mg/l) of BBG111 and its derivative strains in Landy medium with ammonium sulphate at 37°C in Biolector with shaking frequency of 1100 rpm. Results are expressed as mean value with standard deviation	93

Figure 37 Surfactin production (mg/l) of BBG258 and its derivative strains in Landy medium with ammonium sulphate at 37°C in Biolector with shaking frequency of 1100 rpm. Results are expressed as mean value with standard deviation	94
Figure 38 Specific surfactin yield (mg/g DW) of BBG111 and its derivative strains in Landy medium with ammonium sulphate. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test.....	95
Figure 39 Specific surfactin yield (mg/g DW) of BBG258 and its derivative strains in Landy medium with ammonium sulphate. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test.....	96
Figure 40 Surfactin production (mg/l) of BBG111 and its derivative strains in TSS medium supplemented with 16 amino acids at 37°C in Biolector [®] with shaking frequency of 1100 rpm. Results are expressed as mean value with standard deviation.....	103
Figure 41 Specific surfactin yield (mg/g DW) in TSS medium supplemented with 16 amino acids with BBG111 and its derivative strains. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test	104
Figure 42 Comparative analysis on the basis of specific surfactin yield of BBG251 and BBG254 in Landy medium with ammonium sulphate and TSS medium supplemented with 16 amino acids. Results are expressed as mean value with standard deviation	105
Figure 43 The role of volumetric oxygen transfer co-efficient (k_{La}) in determining the maximum specific surfactin yield of BBG258 and BBG260 Results are obtained in shake flask and in Biolector [®] at 37°C and using Landy medium.....	109
Figure 44 Surfactin production (mg/l) by BBG258 and its derivative strains in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation	111
Figure 45 Specific surfactin yield obtained with BBG258 and its derivative strains in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test ..	112
Figure 46 Surfactin production (mg/l) by BBG276 and BBG278 in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation	114
Figure 47 Specific surfactin yield obtained with BBG276 and BBG278 in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation	115
Figure 48 Surfactin production (mg/l) of BBG258 and its derivative strains in classical Landy medium at 6 hour and 8 hour, respectively in Biolector [®] at 37°C with shaking frequency 1100 rpm. Results are expressed as mean value with standard deviation	121

Figure 49 Surfactin production (mg/l) of BBG258 and its derivative strains in AMBER medium at 6 hour and 8 hour, respectively in Biolector [®] at 37°C with shaking frequency 1100 rpm. Results are expressed as mean value with standard deviation	122
Figure 50 Specific surfactin yield obtained with BBG258 and its derivative strains in classical Landy medium at 6 hour and 8 hour respectively. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test.....	123
Figure 51 Specific surfactin yield obtained with BBG258 and its derivative strains in AMBER medium at 6 hour and 8 hour, respectively. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test	124
Figure 52 Regulation of glycolytic pathway and genes marked in red were deleted to analyze their roles in surfactin production	126
Figure 53 Concentration expressed as percentage of each fatty acid chain length in surfactin sample of BBG258 and BBG261	132
Figure 54 Time-resolved analysis of overflow metabolites and its associated metabolic pathways. Bold lines indicate direct conversion, whereas metabolites in boxes indicate the intermediates. Results are expressed as mean value with standard deviation	134
Figure 55 Extracellular concentration of leucine and its degradation product for BBG258 and BBG261, while dashed lines indicated conversion to respective fatty acid and the gene in red was deleted. Results are expressed as mean value with standard deviation.....	136
Figure 56 Extracellular concentration of valine and its degradation product for BBG258 and BBG261, dashed lines indicates conversion to respective fatty acid and the gene in red was deleted	137
Figure 57 Surfactin production (mg/l) by BBG258 and BBG257 in Landy medium with phosphate buffer at 37°C with 250 rpm in shake flask condition. Results are expressed as mean value with standard deviation	143
Figure 58 Specific surfactin yield obtained with BBG258 and BBG257 in Landy medium with phosphate buffer. Results are expressed as mean value with standard deviation	143
Figure 59 Surfactin production (mg/l) by BBG276 and BBG275 in Landy medium with phosphate buffer at 37°C with 250 rpm in shake flask condition. Results are expressed as mean value with standard deviation	145
Figure 60 Specific surfactin yield obtained with BBG258 and BBG257 in Landy medium with phosphate buffer. Results are expressed as mean value with standard deviation	146
Figure 61 Overview of approaches to determine the role of precursors for the enhancement of surfactin in <i>B. subtilis</i>	150

LIST OF TABLES

Table 1 An overview of enzymes and actives molecules along with their corresponding genes responsible in <i>B. subtilis</i>	8
Table 2 Substrate that enhance the synthesis of specific fatty acids in <i>B. subtilis</i> (Kaneda, 1977).....	28
Table 3 Various isoforms of surfactin based on its peptidic moiety along with surfactin (S) native structure	35
Table 4 Different plasmids and their genotype used in this study.....	64
Table 5 Different <i>E.coli</i> strains and their genotype used in this study.....	64
Table 6 Primers used for deletion mutant construction.....	70
Table 7 Primers used for insertion mutant construction.....	73
Table 8 Solutions related to flux where no knockouts are required	86
Table 9 Single knockout predictions for leucine overproduction.....	87
Table 10 Deletion mutants derived from <i>B. subtilis</i> 168 strain based on prediction results of the branched chain amino acid reaction network	89
Table 11 Deletion mutants derived from <i>B. subtilis</i> BSB1 strain based on prediction results of the branched chain amino acid reaction network	90
Table 12 Specific growth rate of all mutant strains derived from BBG111 and BBG258 in Landy medium with ammonium sulphate	92
Table 13 Specific growth rate of BBG111 and its derivative strains in Landy medium with ammonium sulphate and TSS medium supplemented with 16 amino acids	102
Table 14 Specific growth rate of BBG258 and BBG260 in different aeration conditions	107
Table 15 BBG258 and its derivative strains used in this study	110
Table 16 Specific growth rate of BBG258 and its derivative strains in different media ..	110

Table 17 Genotype of BBG276 and BBG278 strain with a constitutive promoter.....	113
Table 18 Specific growth rate of BBG276 and BBG278 in Landy-phosphate buffer medium	114
Table 19 Deletion mutants obtained with <i>Bacillus subtilis</i> BSB1 strain with control based on glycolytic pathway.....	117
Table 20 Integration of <i>sfp</i> gene in the glycolytic mutants obtained with <i>Bacillus subtilis</i> BSB1 strain.....	118
Table 21 Specific growth rate of glycolytic mutant strains and BBG258 in Landy medium and Amber medium	119
Table 22 The mutants obtained with the deletion of <i>yvbW</i> along with the replacement of native promoter with constitutive promoter in <i>B. subtilis</i> strains.....	141
Table 23 Specific growth rate of BBG258 and BBG257 in modified Landy medium buffered with phosphate buffer.....	142
Table 24 Specific growth rate of BBG258 and BBG257 in modified Landy medium buffered with phosphate buffer.....	144

LIST OF ABBREVIATIONS

Microorganisms :

B. subtilis : *Bacillus subtilis*

E. coli: *Escherichia coli*

S. aureus : *Staphylococcus aureus*

Chemicals:

EDTA : Ethylenediametetraacetic acid

MOPS : 3-Morpholinopropane-1-sulfonic acid

X-Gal: 5-bromo-4-chloro-3-indolyl β -galactopyranoside

Metabolites:

Glc : Glucose

BCAA: Branched chain amino acid

Ile : Isoleucine

Val : Valine

Leu : Leucine

Thr : Threonine

Acyl-CoA: Acyl Coenzyme A

Akb : L-2-amino-acetoacetate

Glu: Glutamate

GTP : Guanosine triphosphate

Ket_a : 2-keto-3-methylvalerate

Ket_b : 2-keto-isovalerate

Ket_c : 2-keto-isocaproate

OxoGlu: Oxoglutarate

Pyr: Pyruvate

Methods :

PCR : Polymerase chain reaction

RP-HPLC : Reverse phase – high performance liquid chromatography

NMR: Nuclear magnetic resonance spectroscopy

Units :

OD₆₀₀ : Optical density at 600 nm

h : hour

min : minute

g: gram

mg : milligram

cm : centimetre

l: litre

ml : millilitre

µl : microlitre

M : mole/l

mM : millimole/l

EXTRAS:

NRPS : Non-ribosomal peptide synthesis

DW: Dry weight

GENERAL INTRODUCTION

Motivation *Bacillus subtilis* is a soil bacterium which serves as a model Gram-positive microorganism. This microorganism is considered as “micro-factory” for the production of wide range of industrially beneficial products. The strain *B. subtilis* 168 has been widely studied as well as the entire genome was published in 1997. Depending on the strains, *B. subtilis* can produce various lipopeptides. These peptides are extensively studied for various applications from medicinal to environmental. These biologically synthesized peptides have the potential to replace their chemical counterparts. But the main limitation of its application is its low-level of production, which currently limits their industrial application. Various studies have been carried out to enhance the production of lipopeptides, such as genetic engineering of the strains or bioprocess development for their production and purification. Previous work carried-out in the Research Institute for Food and Biotechnology – Charles Viollette from the University of Lille, Sciences and Technologies, revealed that there might be lack of precursor molecules for their mechanism of biosynthesis, which can have a role in improving the production further. Thus, an intensive and multi-disciplinary approach was required to increase the production of lipopeptides through the enhancement of its precursor concentration. This project was a part of multidisciplinary European project AMBER.

Objective The main objectives of this work were:

1. To analyze the precursor metabolism and define a strategy of gene knockout using computational modelling to increase their intracellular concentration.
2. To optimize the precursor biosynthesis of lipopeptide (surfactin) to increase its production in *B. subtilis*.
3. To analyze the impact of environmental conditions on the surfactin production in the different genetic engineered mutant strains of *B. subtilis*

CHAPTER 1: LITERATURE REVIEW

1. *BACILLUS SUBTILIS*: IN A NUTSHELL

1.1 *Bacillus subtilis* and its genome

Bacillus subtilis is the best-characterized Gram-positive bacteria. It's a rod shaped, spore forming and an aerobic bacterium which is profoundly found in the upper layers of the soil and in association with plants. Under aerobic conditions, it can differentiate into heat-resistant spores. In addition, it can utilize various carbon sources which include many plant-derived molecules. They are widely used for the synthesis of broad range of agricultural, pharmaceutical and chemical compounds, which includes amino acids, antibiotics and biosurfactants. *B. subtilis* has a simple and competent gene transfer system which has led to the construction of genetic map of its genome.

In 1997, Kunst et al. published the entire genome of *B. subtilis* with a genome of 4 million base pairs with an average G+C ratio of 43%, consisting of 4,100 protein-coding genes and later in 2009, the sequence was updated (Barbe et al., 2009). *B. subtilis* CDS can be divided into three different classes. Class I includes mainly genes of sporulation, Class II consists of genes expressed during exponential growth conditions i.e. genes related to transcription and translation machineries, core intermediary metabolism, stress proteins, and one-third of genes of unknown function. Lastly, Class 3 includes genes of unidentified function as well as this class has enriched A+T residues codon (Hénaut et al., 1996). It has been found that ATG as start codon for 78% genes, TTG for 13% and 9% with GTG in comparison to 85%, 3% and 14% for *E.coli* (Blattner et al., 1997).

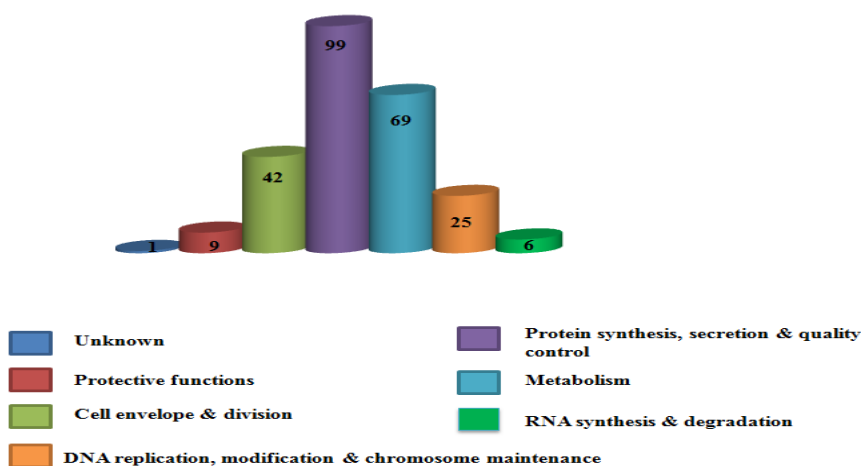


Figure 1 Distribution of essential genes in *B. subtilis* 168 genome

Essential genes can be described as those genes, which are when inactivated under specific condition (LB as nutrient source and 37°C) affect the viability of the cells. Previously, Kobayashi et al. (2003) suggested 271 essential genes but recently Commichau et al. (2013) found out that 31 genes out 271 genes were non-essential. But as per recent updates available in subtiwiki, 251 essential genes are listed. The essential genes have been divided into various domains i.e. protein synthesis, secretion and quality control; metabolism; Cell envelope and division; DNA replication, modification and chromosome maintenance; RNA synthesis and degradation; protective functions and unknown represented in Figure 1.

1.2 The complex regulation of the lipopetptide synthesis

Two-component signal-transduction pathways comprise a sensor protein kinase and a response regulator. In *B. subtilis* 34 genes encode response regulator, largely having neighboring genes encoding histidine kinases. The ComP/ComA two component system is required for the expression of various peptide pheromones, which are involved in various signal transduction pathways (Lazazzera et al., 1999). It has an important role in competence development and regulation of the secondary metabolites biosynthesis, as the lipopeptide surfactin (Hamoen et al., 2003).

B. subtilis can develop a physiological condition which makes them to uptake exogenous DNA in macromolecular form under defined growth condition, this process is known as competence. Hamoen et al (2003) suggested five genetic loci involved for the transport of DNA: *comC*, *comE*, *comF*, *comG* and *nucA*. After transport into the cell, recombination proteins i.e. RecA and DNA helicase AddAB help in integration of the foreign DNA into the genome (Hajjema et al., 1996). ComK is a major candidate for the development competence as all the above genes are regulated by this competence factor (van Sinderen et al., 1995). The crucial step for competence development is the activation of *comK* promoter by ComK (Hamoen et al., 2003). Sporulation gene *spoOA* can also regulate competence as phosphorylated *spoOA* (*spoOA* ~P) can inhibit the expression of *abrB*, which is a potential inhibitor of *comK* expression (Green et al., 1991). Besides inhibition by *abrB*, based on GTP levels as well as concentration of branched chain amino acids *codY* can bind to the RNA polymerase binding site of *comK* promoter, thus inhibiting its expression (Serror and Sonenshein, 1996). Various competence stimulating factors have

direct or indirect effect on the positive expression of *comK*. As it can be seen from Figure 2, expression of *comQ* is required for post-translational modification and secretion of *comX*, which is the main competence stimulating factor. ComX acts as an environmental signal for the phosphorylation of ComA by its protein kinase partner ComP (Solomon et al., 1995). Phosphorylated ComA (ComA~P) is required for the expression of *srfA* operon (Hahn and Dubnau, 1991). It has been found that this operon positively regulates *comK* through *comS*, which is located with the *srfA* operon but out of its ORF (Hamoen et al., 1995).

There are several *phr* genes (7 genes), which encode regulatory peptides that might have a role in quorum sensing mechanism (Perego et al., 1996). PhrC, one of the pheromone which forms another half of competence stimulating factor can also induce competence through *srfA* transcription via ComP and ComA activation. An oligopeptide permease, SpoOK senses this pheromone and facilitates its entry (Lazazzera et al., 1997). PhrC can inhibit the activity RapC. This protein can dephosphorylate ComA~P, thus reducing the expression of *srfA* and finally hindering competence development (Solomon et al., 1996). *B. subtilis* during transition state expresses various extracellular proteases and other degrading enzymes, transport function and chemotaxis to allow maximum utilization of nutrients as well as it excretes various antimicrobial compounds to defeat competing microbial species (Phillips and Strauch, 2002). There are various transition state regulators which inhibit the expression of genes that are only needed during stationary phase (Strauch et al., 1993). This transition state regulator includes AbrB, ScoC, SinR, CodY and Abh. AbrB, ScoC and SinR overexpression can restrain sporulation (Gaur et al., 1986) while AbrB is involved in the production and resistance to various antibiotics (Hahn et al., 1995). Both SpoOA and AbrB regulate the expression of *scoC* (Perego and Hoch, 1988). SinR represses the genes (*spoIIA*, *E* and *G*) responsible for sporulation, it is a global regulator for degrading enzymes synthesis, motility, competence and sporulation (Gaur et al., 1991). CodY is activated in the presence of amino acids (valine and isoleucine) and GTP; it represses competence and dipeptide permeases as well as genes involved in the utilization of alternate carbon and nitrogen sources (Serror and Sonenshein, 1996).

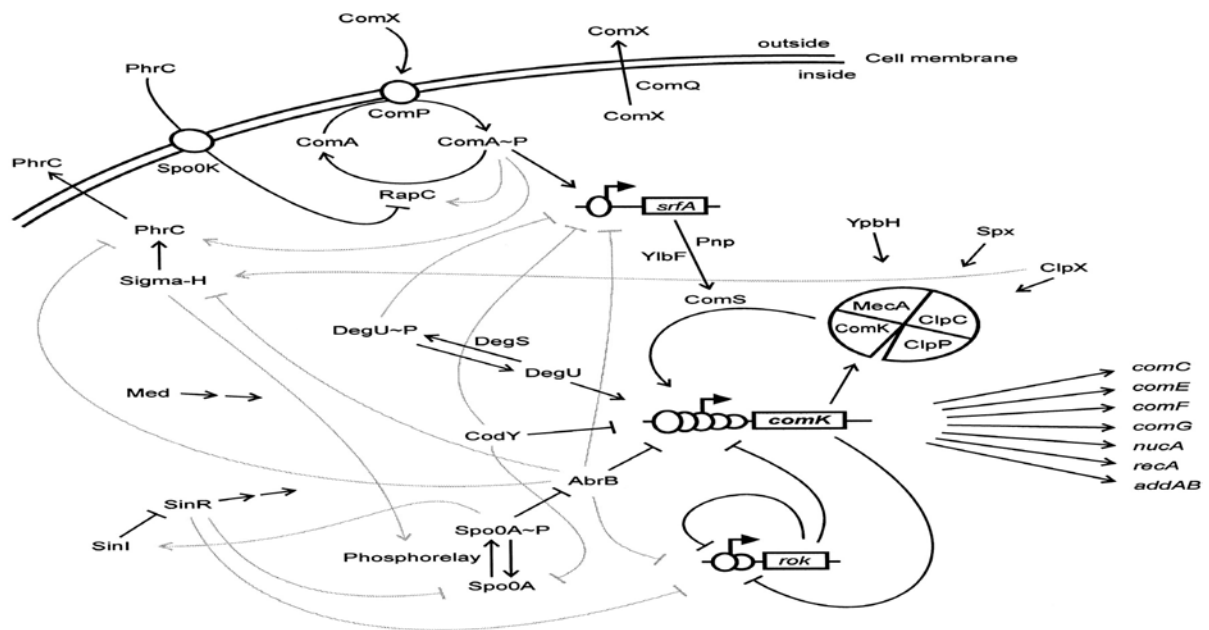


Figure 2 Overview of regulatory pathways affecting competence development as well as positive and negative regulation of *comK* and related genes are indicated by arrows and T-bars (Hamoen et al., 2003)

The most abundant class of proteins found in *B. subtilis* are ABC transporters as Gram-positive bacteria have an envelope comprising single membrane. There are 77 such transporters encoded in the genome (Kunst and al., 1997).

1.3 Industrial applications

Bacillus species have been the “protagonists” in applied microbiology for industrial applications for the last 20 years and more. In 1982, *B. subtilis* (natto) was first used for the solid state fermentation of soyabeans in Japan. They have been drawing attention towards industrial application due to the following characteristics: high growth rates, ability to secrete proteins and various important compounds into extracellular medium and generally recognized as safe (GRAS) organisms for food and drug administration. The availability of the genomic information related to various pathways have facilitated reconstruction of this pathway through gene modification, as role expression studies can be carried out for the overproduction of various industrially important compounds along with

the characteristics of development of competent state to uptake foreign DNA and sporulation. There has been extensive proteome analysis providing a panoramic view of *B. subtilis* proteome and its ability to adopt environmental changes (Becher et al., 2011). There are various genes which are responsible for the production of lipid, carbohydrate and protein degrading enzymes as well as responsible for secretory function, thus creating an industrial interest. *B. subtilis* secreted proteins have N-terminal signal peptides, which are important feature for transport from the cytoplasm (Dalbey et al., 2012) and the best studied translocation pathway for secretory proteins are secretory (Sec) pathway and twin-arginine translocation pathway (Antelmann et al., 2003). To enhance the biotechnological applications, large scale genome minimization studies are being carried out, strain (MBG874) lacks 874kb (20.7%) of the *B. subtilis* genome and *B. subtilis* $\Delta 6$ lacks 376 genes at seven different locations in reference to *B. subtilis* 168 (Reuß et al., 2016). Industrial enzymes market will be worth 6.2 billion \$US by 2020, *B. subtilis* can be the major workhorse for this market. Its importance comes from its ability to produce these substances at gram per litre concentrations. *B. subtilis* produces α -amylase, which cleaves internal α -1, 4-linkages in an endo fashion and it is used in food, fermentation, textile and paper industries (Pandey et al., 2000). The enzyme β -glucanase is used for enzymatic modification of barley. It can produce various alkaline proteases which have pH optima in the range of 9-12, is thermostable as well as can be used as commercial detergents (Gupta et al., 2002). Two novel phytases from *B. subtilis* were cloned and biochemically characterized (Tye et al., 2002).

Braaz et al. (2002) found that *B. subtilis* was an efficient host for the overexpression of poly (3-hydroxybutyrate) depolymerase from *P. lemoignei*. *B. subtilis* can be a great host for the production of heterologous proteins as it has the potential to secrete proteins into extracellular medium but it requires various genetic signals for effective transcription and translation of foreign genes (Braaz et al., 2002). Production of heterologous eukaryotic proteins in *B. subtilis* is problematic due to the presence of large number of extracellular proteases (Wang et al., 1988). About 4% of the *B. subtilis* genomes code for large multifunctional enzymes which include genes for loci i.e. Srf, Pps and Pks. Due to presence of these genes, it can produce compounds with antibiotic activity. Bacitracin gene cluster was transferred from *B. licheniformis* to *B. subtilis*; it was able to produce the antibiotic in high levels (Eppelmann et al., 2001). Production of riboflavin at the

concentration of 30 g/l in 3-day fermentation was obtained using mutant strain of *B. subtilis* (Dauner et al., 2002).

Table 1 An overview of enzymes and active molecules along with their corresponding genes in *B. subtilis*

GENE	PRODUCT
<i>amyE</i>	α -Amylase
<i>aprE</i>	Subtilisins
<i>bglH</i>	β -Glucosidase
<i>bglS</i>	B-Glucanase
<i>csn</i>	Chitosanase
<i>ganA</i>	β -Galactosidase
<i>lipA</i>	Lipase A
<i>nprB</i>	Extracellular neutral protease B
<i>pel</i>	Pectate lyase C
<i>ppsA</i>	Fengycin
<i>urfA</i>	Surfactin
<i>xynC</i>	Endo-xylanase
<i>ribA</i>	Riboflavin
<i>bioA</i>	Biotin

B. subtilis spores can also be used as an agent for vaccine delivery (Duc et al., 2003). System and synthetic biology approach along with fundamental and applied approach can convert *B. subtilis* into micron-scale incomparable factories.

It can be seen that *B. subtilis* genome is properly studied to be used for genetic modification for the synthesis of various molecules. In the next section, the focus is on non-ribosomal peptide synthesis (NRPS) mechanism in *B. subtilis*, which is essential for the production of particular class of industrially important molecules, such as lipopeptidic biosurfactants.

2. INSIGHTS INTO UNIQUE NON-RIBOSOMAL PEPTIDE SYNTHESIS

2.1 Introduction

In 1928, Alexander Fleming returned to his laboratory after a month and saw his culture of *Staphylococcus aureus* had become contaminated with a mold. He was unaware of the fact that this incident will act as a catalyst in the world of medicine with the discovery of first antibiotic (which he called “mold juice”, later renamed as penicillin) and the mold was *Penicillin notatum*. Penicillin is a peptide derivative from *Penicillin notatum*. Lipmann and his co-workers (1971) communicated enzymatic synthesis of tyrocidine and gramicidin S from *Bacillus sp.* It was reported that their synthesis was independent on nucleic acid and was carried out by a large multifunctional enzyme complex analogous to fatty acid synthetases (Lipmann et al., 1971).

Non-Ribosomal Peptide Synthetase (NRPS) is a multifunctional, multienzyme complex found in various bacteria, fungi and lower eukaryotes. In comparison to conventional ribosomal synthesis, it does not require mRNA or ribosome for the synthesis of the peptides. Non ribosomal peptides (NRPs) display more than 500 different monomers and can be linear, cyclic or branched in structure. NRPs can be classified into lipopeptides, chromopeptides, glycopeptides etc. There is another class of multifunctional enzymes known as PKS (Poly Ketide Synthetases). The difference between PKSs and NRPs is that, the former uses acetate and propionate as building blocks while the latter uses monomers (Mootz et al., 2002). NRPs represent a large class of bioactive compounds with distinctive structural features like D- and L- amino acids; heterocyclic elements which can be methylated or glycosylated. Non-ribosomal peptides are generally produced by microorganisms present in soil e.g. *Bacilli*, *Actinomycetes* or *Pseudomonas*. Although, it has been found that marine microorganisms can be a resource for these multifunctional secondary metabolites (Faulkner, 1998). These peptides display various functions i.e. antibiotic activity (Tyrocidine A; Bacitracin A; Vancomycin), biosurfactant (Surfactin, Iturin), antifungal (Iturin, Fengycin), elicitor (Surfactin, Fengycin), anti-inflammatory (Surfactin) and act as immunosuppressive agent (Cyclosporin) as well as can function as

siderophore (Bacillibactin; Enterobactin; Vibriobactin)(Marahiel, 2009). They can also be divided into classes based on their structural conformation: linear product (vancomycin, etc.) and cyclic product (surfactin, cyclosporine, etc.) following intramolecular nucleophilic reaction.

NRPS is organized into various modules which forms the template for the synthesis of various linear and cyclic peptides. Modules are further divided into domains, which are responsible for specific catalytic activity during synthesis (Felnagle et al., 2008). Each module determines the nature of the monomer and inclusion of the monomer in the growing peptide (Caboche et al., 2008) whereas the domains are responsible for substrate binding, activation, modification, elongation and release of the complete peptide (Sieber, 2005). The length of the final peptide is generally determined by the size of the NRPS. Studies have revealed the presence of linker regions (~15 amino acids) between domains in a randomly distributed fashion. These regions being flexible in nature can be used for synthesizing hybrid NRPS template without disturbing the enzymatic integrity (Weber et al., 2000). A database consisting of all non-ribosomal peptides have been created called NORINE (Caboche et al., 2008). The biological activity of NRPS depends on their structure, which is defined by the peptide sequence and the length and isomery of the fatty acid chain for the lipopeptidic compounds. Sometimes, the peptide products can become problematic for the producer strain, so they acquire several defensive mechanism i.e. modification of the target region within the strain, temporary inactivation of the product and efflux pumps to escape the toxic nature (Crosa and Walsh, 2002).

2.2 Primary domains

The elongation module (C-A-PCP) is made of at least three essential domains i.e. Adenylation (A) domain, Peptidyl Carrier Protein (PCP) domain and Condensation (C) domain, for the selection and activation of the monomer, covalent binding to 4'-phosphopantetheine (PPan) and peptide bond formation, respectively. At the end, the final peptide release is catalyzed by thioesterase (TE) domain. These four domains make the primary domains of Non-Ribosomal Peptide Synthetase.

2.2.1 Adenylation (A) domain

Adenylation (A) domain is responsible for the selection and conversion of the monomer (amino or carboxylic acid). Upon binding, the domain converts the substrate into aminoacyl adenylate and release PPi in the presence of ATP and Mg²⁺ (Figure 3). The amino acids present in the core motifs are responsible for the ATP binding and hydrolysis as well as adenylation intermediate formation (Stachelhaus et al., 1999). The Adenylation domain is composed of ~500 amino acids (Mootz and Marahiel, 1997).

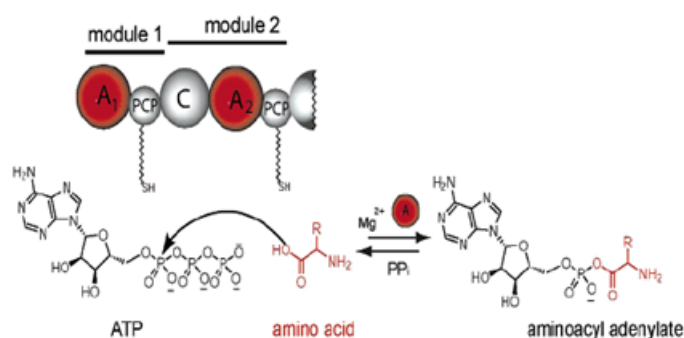


Figure 3 Amino acid recognition and activation by ATP hydrolysis by Adenylation (A) domain (Sieber and Marahiel, 2005)

This domain belongs to the superfamily known as adenylate-forming enzymes. Its functional requirement of ATP is similar to aminoacyl-tRNA-synthetases (Sieber and Marahiel, 2005). The hydrophobic active site is between the large size N-terminal subdomain and C-terminal subdomain, which is small and flexible. The latter subdomain adopts different positions relative to N-terminal subdomain to facilitate the initial catalytic reaction. This orientation is favored by a rotation of 140° during open state conformation (substrate loading) relative to the closed state conformation (Adenylation reactions) (Yonus et al., 2008).

2.2.2 Peptidyl carrier protein (PCP) domain

PCP domains consist of PPan (4'-phosphopantetheine) prosthetic arm by the activity of a specific PPan-transferase (Mootz et al., 2001), which is posttranscriptionally attached onto

PCP domains to convert apo-PCP to holo-PCP. The prosthetic arm is attached through a highly conserved serine residue (Marahiel, 2009). PCP domain can have three interconverting conformations (A, A/H and H) based on the chemical state. PCP domain binding with 4'-phosphopantetheine at the serine residue converts A/H (acyl-holo-PCP) state into H (holo-PCP) state. Mutation of active site serine residue to alanine can convert H (holo-PCP) to A (apo-PCP) state, which can generate a defeat in protein partner recognition (Koglin et al., 2006). NMR studies reveal PCP domain consists of a large loop region (contains the serine residue which is attached to PPan) between two helical regions. The PCP cofactor was modified by Sfp-Glu151. This residue of Sfp deprotonates the serine-hydroxyl group of PCP domain (Figure 4) and the subsequent nucleophilic reaction on the β -phosphate of CoA leads to the formation of holo-PCP and 3',5'-ADP (Mofid et al., 2004). PCP domain has a four-helix bundle structure and Sfp binds within an extended loop between the first two helices (Weber et al., 2000).

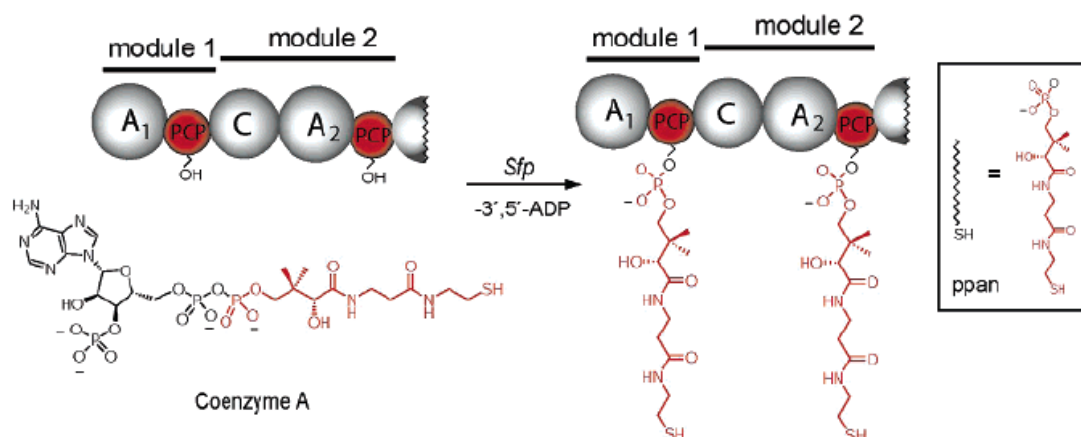


Figure 4 Role of product for the attachment of prosthetic arm (Sieber and Marahiel, 2005)

Sfp monomers have a pseudo-2-fold symmetry and coenzyme A binds within a pocket formed by Sfp bent conformation (Sieber and Marahiel, 2005). It was observed that *sfp* gene product can efficiently phosphopantetheinylate not only apo-PCP (NRPS templates) but also acyl carrier proteins (fatty acid) of polyketide synthases (Kealey et al., 1998). Phosphopantetheine (PPan) transferase expressed by *sfp* validated the findings that PCP domain's protein-protein interaction depends on its conformation state (Lai et al., 2006).

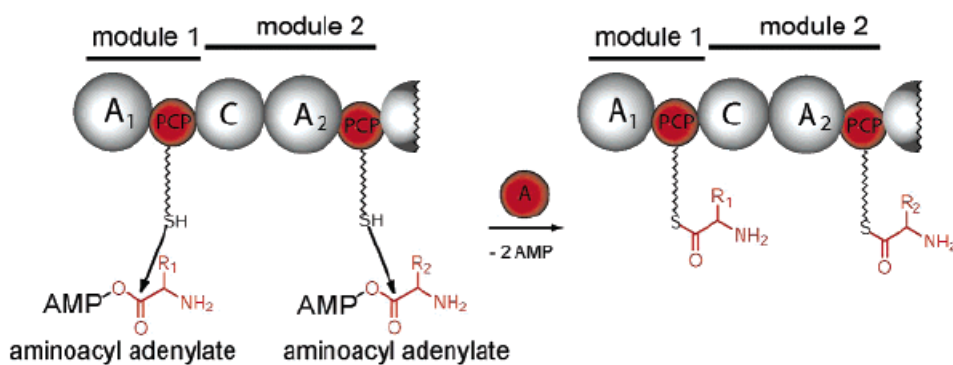


Figure 5 Activated amino-acyl adenylate forming a covalent bond with the free thiol of PCP domain (Sieber and Marahiel, 2005)

Initiation module is formed by both adenylation (A) and PCP domain as they are required for the initial activation and covalent bond formation for the subsequent peptide synthesis. The reactive intermediate (amino-acyl adenylate) from A domain is then transferred to the thiol moiety of holo-PCP domain (Figure 5) by a nucleophilic reaction between the free thiol-PPan cofactor and amino acid -O- AMP, with the formation of a thioester (Sieber and Marahiel, 2005). The holo-PCP domain composed of 80-100 amino acids. It acts as a swinging arm for the transfer of thioester bound substrate and elongation intermediate from A domain to C domain.

2.2.3 Condensation (C) domain

Condensation domain is responsible for the formation of peptide bond between the PCP intermediates. This domain is composed of ~450 amino acids. The peptide bond is formed by a nucleophilic reaction between the acceptor site α -amino group of the aminoacyl-S-PCP and peptidyl-S-PCP substrate in the donor site (Figure 6) i.e. PCP-bound acceptor amino acid downstream and upstream PCP activated donor substrate (Stachelhaus et al., 1998).

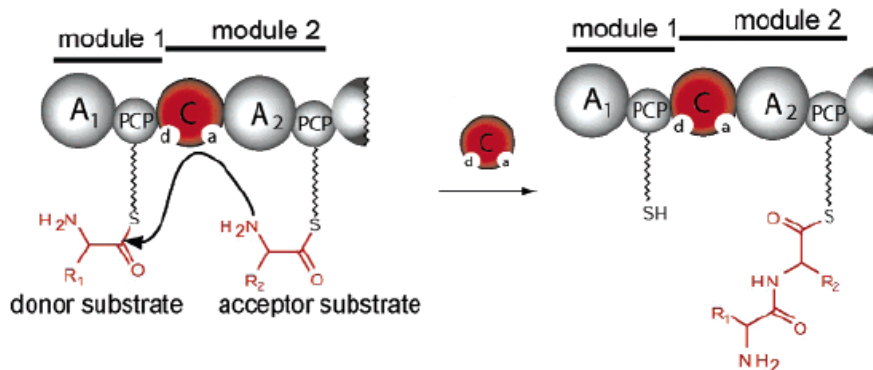


Figure 6 Peptide bond formation between acceptor (a) substrate and donor (d) substrate is catalyzed by C-domain (Sieber and Marahiel, 2005)

In vitro analysis revealed C-domain has a high substrate specificity at the acceptor in comparison to the donor site (Belshaw et al., 1999). The C-domain has a V-shaped conformation for the binding of two PCP substrates and an active site having a conserved core motif located in close proximity to the binding site (Samel et al., 2007). Studies on the active site pK value revealed that the peptide bond formation depends on electrostatic interactions. The C-domain has been replaced by heterocyclic (Cy) domain in synthesis of vibractin, as it can perform similar functions like peptide elongation and heterocyclization in a similar way as C-domain (Marahiel, 2009).

2.2.4 Thioesterase (TE) domain

Thioesterase (TE) domain is composed of ~280 amino acids (Marahiel, 2009). Its structure consists of a fold of α/β hydrolases and a lid region which covers the active site consisting of a conserved catalytic triad (Ser, His and Asp). The lid region has both open and closed conformations for preventing the active site from water (Samel et al., 2006). It is only found in the termination module of NRPS.

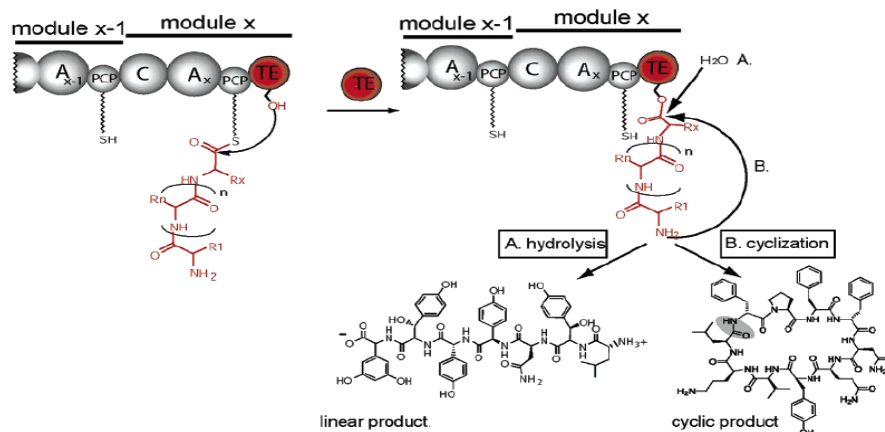


Figure 7 Role of TE domain for peptide release (Sieber and Marahiel, 2005)

This domain is responsible for the termination of synthesis (Figure 7) by two-step reaction: i) Transferring the final peptide to the active serine residue of the terminal PCP domain of the NRPS assembly (with the formation of oxoester linkage from thioester linkage); ii) producing peptidyl-acyl-enzyme intermediate (acyl-O-TE) following dissociation by an internal nucleophilic reaction (intramolecular region and stereoselective macrocyclization) or hydrolysis (linear product) (Trauger et al., 2000). TE domains are responsible for cyclization of diverse nonribosomal peptides. The diverse nature can be distributed on the basis of two predominant different cyclization strategies: 1 lactone (Surfactin, Enniatin, Fengycin, Syringomycin etc.); 2 lactam (Iturin A, Mycosubtilin, Tyrocidine, Cyclosporin A, Gramicidin S etc.). Lactone and lactam cyclization strategies can be differentiated on the basis of processing, while in the former ketone is reduced to hydroxyl group and in the latter, it is converted to amino group by head to tail condensation (Sieber and Marahiel, 2005). An alternative pathway of termination can take place by the aid of reductase (R) domain in the terminal by the reduction of peptidyl-S-PCP final product (Kopp and Marahiel, 2007). This domain has been discussed later in the chapter.

Analysis of TE domain of SrfA towards substrate tolerance, showed the significance of both N- and C- terminal monomers forming the peptide. This result revealed that leucine (6 and 7 position) and glutamic acid (1st position) in the peptide moiety of surfactin are essential monomers for TE domain for substrate recognition (Trauger et al., 2001).

In addition to natural synthesis of non-ribosomal peptides, chemoenzymatic approaches have also been carried out utilizing the versatile nature of TE domains. In order to carry

out the synthesis, TE domains were provided with substrate attached to PCP domain or an artificial solid support or as soluble thioester intermediates (Sieber and Marahiel, 2005).

2.3 Secondary domains

There are five types of secondary domains, which include epimerization (E) domains, N-methyltransferase (Mt) domain, Formylation (F) domains, and oxidation (Ox) and reduction (Red) domains.

2.3.1 Epimerization (E) domain

This domain transforms the L-isomers (aminoacyl and peptidyl thioesters) attached to PCP domain into corresponding D-isomers (Marahiel, 2009). In few cases, epimerization can also be catalyzed by an external reaction e.g. in the synthesis of cyclosporin, A domain includes D-Ala provided by an external racemase (Hoffmann et al., 1994).

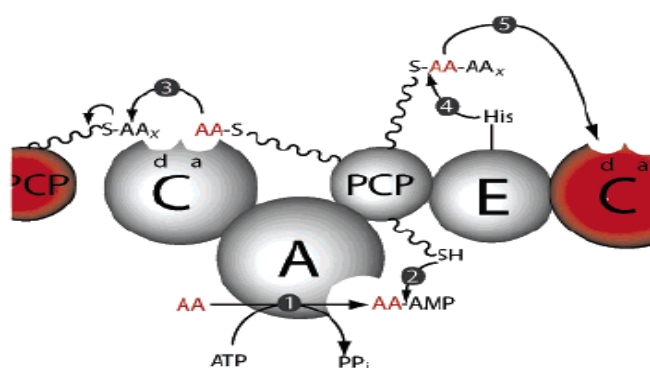


Figure 8 Role of epimerization (E) domain during elongation (Sieber and Marahiel, 2005)

As shown in Figure 8, followings steps takes place during elongation of a particular peptide: 1 the substrate is activated by adenylation; 2 loading of reactive intermediate into PCP domain; 3 binding into acceptor site of C-domain and subsequent covalent bond formation; 4 movement to E domain and epimerization, 5 D-isomer binding to the donor site of C-domain and then sequential reaction (Sieber and Marahiel, 2005).

2.3.2 N-methyltransferase (N-Mt) & C-methylation (C-Mt) domain

N-methyltransferase (Mt) domain is placed adjacent to PCP domain and is responsible for transferring an electrophilic methyl group from *S*-adenosylmethionine (SAM) to the α -amino group (Figure 9) of the thioesterified amino acid, peptide, thus forming *S*-adenosylhomocysteine (Schauwecker et al., 2000). N-methylated peptide bonds are present in peptides i.e. cyclosporin, enniatin, actinomycin and pristinamycin (Sieber and Marahiel, 2005).

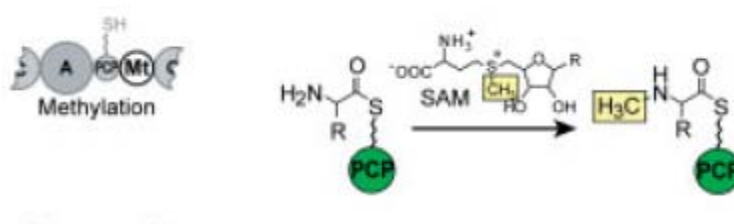


Figure 9 Reaction involving N-methyltransferase (Mt) domain, the yellow box indicates the modification (Marahiel, 2009)

There is another secondary domain which is responsible for methylation at the C-terminal. C-Mt domain catalyzes the C α -atoms of cysteine residue into α -methyl-thiazoline. This domain also depends upon SAM for the methyl group. This domain is accountable for cyclization of the peptide and located at the C-terminal of PCP domain (Marahiel, 2009).

2.3.3 Formylation (F) domains

This domain is responsible for the transfer of formyl group to the peptide chain during initiation (Figure 10) (Marahiel, 2009).

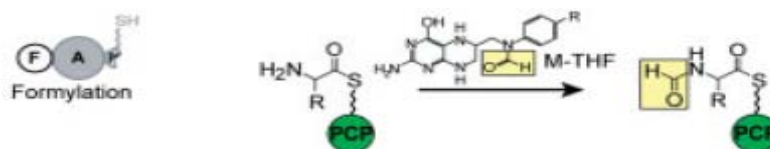


Figure 10 Reaction involving Formylation (F) domain, the yellow box indicates the modification (Marahiel, 2009)

2.3.4 Other modification domains

Oxidation and reduction domains (Figure 11) are incorporated within A-domain C-terminal region. The former in the presence of FMN produces thiazoles and oxazoles, whereas the later produces thiazolidine and oxazolidine rings in the presence of NADPH (Marahiel, 2009). In vancomycin, six of the seven amino acids are nonproteinogenic; this modification can be carried out by the oxidation domain (Vaillancourt et al., 2006).

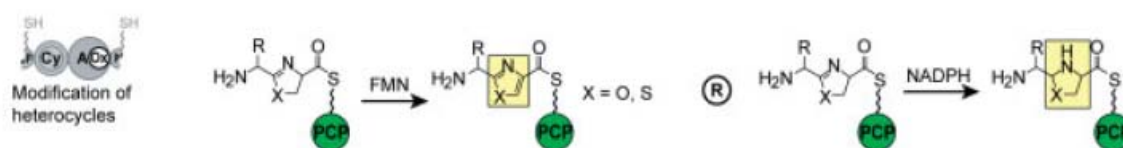


Figure 11 Reaction involving oxidation (Ox) and reduction (R) domains, the yellow box indicates the modification (Marahiel, 2009)

Certain peptides i.e. surfactin and fengycin have fatty acid in their N-terminal. It is assumed that acyl transferase provides the acyl to donor site of C-domain to have this modification (Steller et al., 1999).

2.4 Classification

Based on the interaction between the modules or domains, NRPSs mode of biosynthesis can be classified into three different categories (Mootz et al., 2002):

1. Linear NRPS (type A)
2. Iterative NRPS (type B)
3. Nonlinear NRPS (type C)

2.4.1 Linear Synthesis (type A)

In this mode of biosynthesis; the final peptide contains n amino acids, which is incorporated by n modules with a domain arrangement A-PCP-(C-A-PCP) $_{n-1}$ -TE (Figure 12), while the secondary domains are placed in their respective modules (Henning D. Mootz et al., 2002). The initiation (A-PCP) module lacks (except for some lipopeptide) the C-domain whereas C-A-PCP forms the elongation and TE domain and in some cases modified C-domain (peptides of fungal origin) forms the termination module. In this type of NRPS, the order and number of modules determines the sequence of the final peptide (Keating et al., 2001). Surfactin, cyclosporin, tyrocidine, actinomycin, bacitracin, fengycin, etc are examples of linear NRPSs (Mootz et al., 2002).

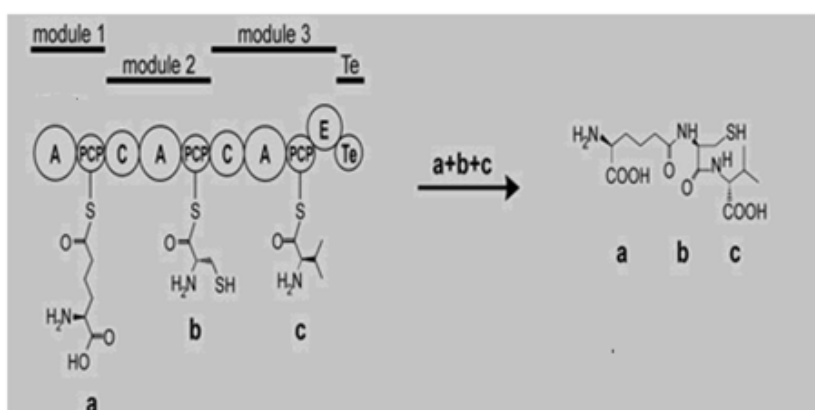


Figure 12 Assembly and mode of synthesis of linear NRPSs (Mootz et al., 2002)

2.4.2 Iterative Synthesis (type B)

In this type of biosynthesis, the modules or domains are used more than once for the synthesis of final peptide. It contains reduced number of modules relative to the sequence of the final peptide. One set of repetition is provided by the initiation module, while the other module is responsible for the cyclization of the product by oligomerization process as per example (Figure 13). The repeated cycles as well as the bond nature between the monomers are controlled by PCP-C didomain and TE domain (Mootz et al., 2002).

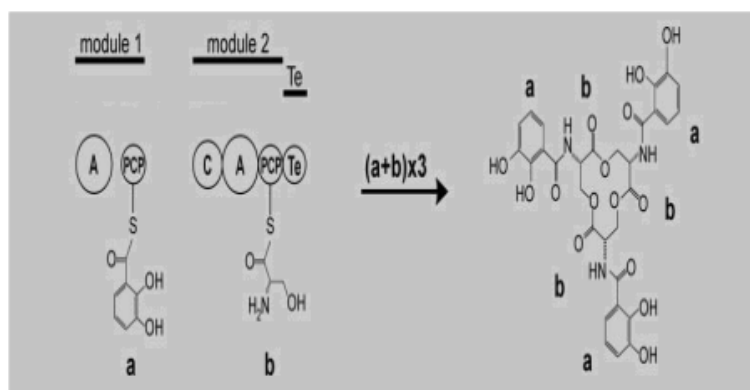


Figure 13 Assembly and mode of synthesis of iterative NRPSs (Mootz et al., 2002)

Examples of this type of synthesis include enterobactin produced by *E.coli* and enniatin produced by *Fusarium scirpi* (Haese et al., 1993).

2.4.3 Nonlinear Synthesis (type C)

The modular arrangement in nonlinear synthesis is different in comparison to linear NRPSs (Figure 14). This structural difference leads to unusual internal conformation. Syringomycin, yersniabactin, vibriobactin, mycobactin, bleomycin etc. are examples of this type of biosynthesis (Mootz et al., 2002).

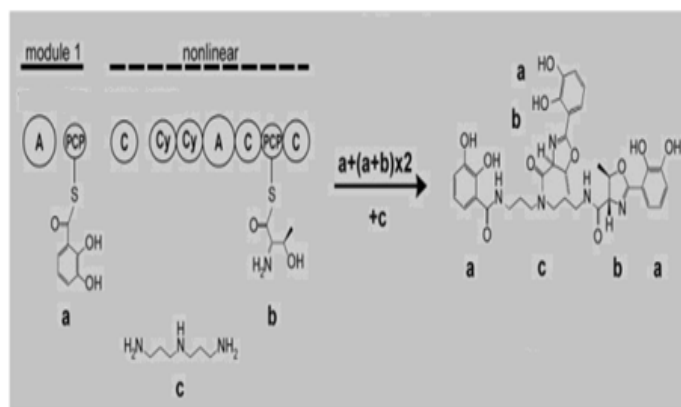


Figure 14 Assembly and mode of synthesis of nonlinear NRPSs (Mootz et al., 2002)

2.5 Precursors of the NRPS mechanism and their regulation mechanisms

B. subtilis can utilize diverse carbon sources. Non-Ribosomal Peptide Synthetase (NRPSs) mechanism requires various precursors for the synthesis of different peptides. The precursors are generally produced by various metabolic pathways involved in the cell. The role of various pathways required for the production of different metabolites have been discussed below.

2.5.1 Glycolytic pathway

B. subtilis displays Embden–Meyerhof–Parnas glycolytic pathway, linked to a functional tricarboxylic acid cycle. The pathway comprises of two sections: the upper section is responsible for the conversion of hexose carbon source to triose and the bottom portion is responsible for pyruvate synthesis using the product from upper section as substrate. There are numerous enzymes involved in the import and conversion of glucose into pyruvate and finally pyruvate is used as a substrate for TCA cycle or other pathways for the production of various metabolites required for metabolism, cell wall synthesis and synthesis of secondary metabolites. A summarized overview of this pathway is provided in Figure 15.

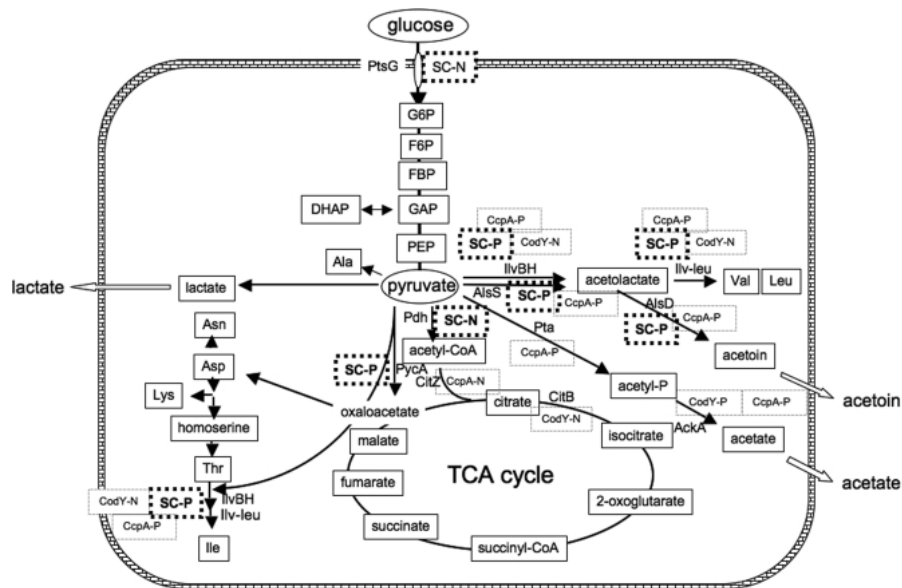


Figure 15 Role of glycolytic pathway and production of various metabolites as a result of this pathway (Tojo et al., 2010)

There are various transcriptional regulators which regulates the glycolytic pathway in addition to enzymes responsible for the conversion of substrates. These regulators include *ccpA*, *ccpN* and *cggR*. The global regulator *ccpA* is activated through binding of HPr protein. The activity of phosphorylated CcpA is enhanced through the binding of glucose-6-phosphate and fructose-1,6-phosphate. CcpA has a positive effect on *gapA* operon (Sonenshein, 2007). There are two transcriptional repressor *ccpN* and *cggR* in the glycolytic pathway. The former is expressed constitutively; it binds in the promoter region and can repress the expression of *gapB*. While the latter is controlled by the concentration of fructose-1,6-bisphosphate and can repress *gapA* operon (Servant et al., 2005).

The regulon *ptsGHI* is the first set of gene clusters involved in the glycolytic pathway. The gene *ptsG* (glucose permease) enables transport of glucose into the cell and phosphorylates glucose to glucose-6-phosphate via HPrK/P (Bifunctional kinase/phosphatase). The activity of HPrK/P is stimulated by ATP and fructose-1,6-bisphosphate. The transfer of phosphoryl group from phosphoenol pyruvate (PEP) to permease is carried out by *ptsH* and *ptsI*. This system only produces glucose-6-phosphate (Ludwig et al., 2001). The RNA binding protein *gltT*, acts as a transcriptional antiterminator, which in the presence of glucose binds stem-loop structure of *ptsGHI* mRNA, thus preventing formation of a

terminator structure and allowing transcription, but in the absence of glucose it cannot bind to *ptsGHI* mRNA and no transcription takes place (Commichau and Stülke, 2008). Glucose-6-phosphate can also be obtained from glucose-1-phosphate by *pgcA* (α -phosphoglucomutase). Glucose-1-phosphate is generally produced by other sugars except glucose (Weart et al., 2007). Glucose-6-phosphate is then converted into fructose-6-phosphate by a reversible reaction carried out by *pgi* (glu-6-phosphate isomerase), which is constitutively expressed and does not depend on the presence of glucose (Macek et al., 2007). The conversion of fructose-1,6-biphosphate to fructose-6-phosphate can be carried out by two types of fructose-1, 6-biphosphatase *fbp* or *glpX* (Fujita and Freese, 1981) (Fujita et al., 1998). AMP and PEP concentrations regulate the expression of *fbp* while only PEP regulates the expression of *glpX* (Jules et al., 2009). Fructose-6-phosphate is converted to fructose-1,6-biphosphate by *pfkA* (phosphofructokinase). The gene *fbaA* (fructose-1,6-biphosphate aldolase) is responsible for the conversion of fructose-1,6-biphosphate to glyceraldehyde-3-phosphate and dihydroxyacetone phosphate. This enzyme is expressed in a constitutive manner. Dihydroxyacetone phosphate is further used for the synthesis of glycerol by *glpKD*, while interconversion of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate by enzyme is expressed by *tpi*. The enzyme encoded by *gapA* (glyceraldehyde-3-phosphate dehydrogenase) is responsible for the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphateglycerate. The backward reaction (1,3-bisphosphateglycerate to glyceraldehyde-3-phosphate) is favored by *gapB* (glyceraldehyde-3-phosphate dehydrogenase 2). The gene *pgk* (Phosphoglycerate kinase) is responsible for the reversible conversion of 1,3-bisphosphateglycerate and 3-phosphateglycerate as well as it requires magnesium and manganese ions as cofactors (Suzuki et al., 1974). The reversible conversion of 3-phosphateglycerate to 2-phosphateglycerate is catalyzed by *pgm* (2,3-biphosphoglycerate-independent phosphoglycerate mutase). The interconversion of 2-phosphateglycerate to phosphoenol pyruvate is catalyzed by *eno* (enolase). It interacts with *pfkA* in vitro and increases the flux in glycolytic pathway (Newman et al., 2012). The final reaction of glycolytic pathway, conversion of phosphoenol pyruvate to pyruvate is carried out by *pykA*.

The essential genes in the glycolytic pathway are *eno* (enolase; responsible for the conversion of 2-phospho-D-glycerate to phosphoenolpyruvate), *gapA* (glyceraldehyde 3-phosphate dehydrogenase; converts D-glyceraldehyde 3-phosphate to 3-phospho-D-glyceroyl phosphate) and *pgm* (phosphoglycerate mutase; conversion of 2-phospho-D-

glycerate to 3-phospho-D-glycerate) as the mutants lacking these genes were unable to grow on LB-medium (Commichau et al., 2013).

Pyruvate produced from the glycolytic pathway along with threonine is directed towards the synthesis of branched chain amino acids i.e. valine, leucine and isoleucine. The metabolism of branched chain amino acids (BCAAs) is described in the following section.

2.5.2 Branched chain amino acid pathway

The branched chain amino acids (BCAA), i.e. leucine (Leu), valine (Val), and isoleucine (Ile) are produced by this metabolic pathway with complex regulatory mechanisms. Threonine (Thr), pyruvate (Pyr) and L-2-amino-acetoacetate (Akb) are used by the cell for the production of leucine, valine, and isoleucine. The *ilv-leu* operon contains the main genes involved in the production of branched chain amino acids. The *ilv-leu* operon is subjected to multiple forms of regulation on its promoter $P_{Ilv-Leu}$, as studied in a series of articles. There are two proteins that down-regulate the activity of $P_{Ilv-Leu}$ by binding to different sites upstream the promoter and inhibiting the RNA polymerase (Belitsky and Sonenshein, 2011). These are CodY (transcriptional pleiotropic regulator, which regulates the transcription of early-stationary-phase genes) and TnrA (nitrogen pleiotropic transcriptional regulator). In addition, the down-regulation of the promoter $P_{Ilv-Leu}$ is also contributed by Leu in terms of ribosome-mediated attenuation mechanism, Tbox (Brinsmade and Sonenshein, 2011). The expression of different genes involved in the production of branched chain amino acids is shown in Figure 16.

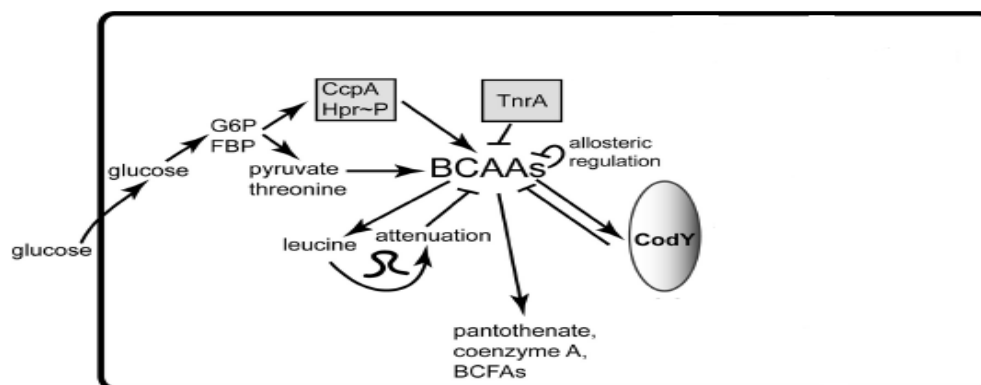


Figure 16 Complex regulation of branched chain amino acid pathway (Brinsmade et al., 2010)

As is shown in Figure 17, the protein IlvA activates the conversion of Thr into Akb. IlvBH (a cluster of IlvB and IlvH), IlvC and IlvD activate the transformation of Akb to Keta. The proteins LeuA and LeuBCD (a cluster of LeuA, LeuB, and LeuC) perform a sequence of reactions justifying the transformation of Ketb to Ketc (Mäder et al., 2012). The genes coding the proteins IlvBH, IlvC, LeuA, and LeuBCD are co-located in the same operon (*ilv-leu*), so they are under the dependence of the same promoter $P_{Ilv-Leu}$ (Tojo et al., 2005). Expression of *ilvA*, *ilvD*, and *ybgE* is dependent on their own promoters and is down-regulated by CodY at the transcriptional level. The expression of *ilvA* is deactivated by both Ile and Val (Mäder et al., 2004a). The transcription of all proteins IlvC, IlvBH, LeuA, and LeuBCD starts through the activation of their promoter $P_{Ilv-Leu}$. The role of various genes in *ilv-leu* operon and others in the production of isoleucine, leucine and valine (ILV) is shown in Figure 17.

The intermediates, 2-keto-3-methylvalerate (Ket_a); 2-keto-isovalerate (Ket_b) and 2-keto-isocaproate (Ket_c) used for the synthesis of isoleucine, valine and leucine are transformed by the protein pairs YwaA and YbgE (Thomaides et al., 2007). The intermediates, Ket_a is produced from Akb, Ket_b from Pyr and Ket_c from Ket_b . Akb can be produced from Thr, coming from the threonine biosynthesis pathway (which is linked to the glycolysis pathway via the aspartate biosynthesis) while Pyr is produced by the glycolysis pathway. The creation of Ile, Val, and Leu from Ket_a , Ket_b and Ket_c is always coupled with a parallel transformation of glutamate (Glu) into 2-oxoglutarate (OxoGlu), and the inverse decompositions are coupled with the inverse transformation of OxoGlu into Glu. The

required Glu for the production of Ile, Val, and Leu, can be synthesized from the glutamate biosynthesis pathway.

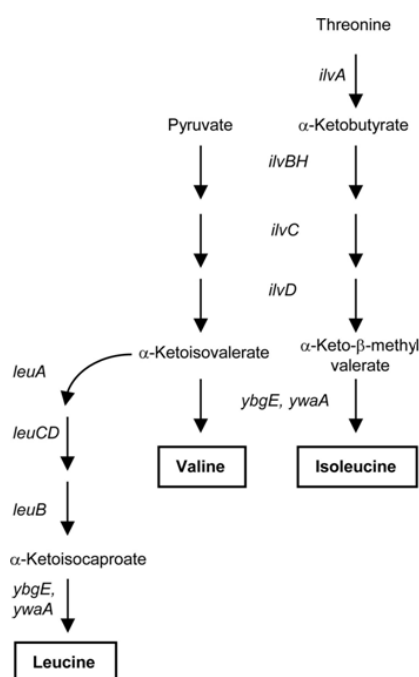


Figure 17 Role of various genes for the production of isoleucine, leucine and valine (Belitsky, 2015)

The intermediates, Ket_a; Ket_b and Ket_c are used for fatty acid synthesis and transformed by protein BkL, this metabolic pathway has been described in the next section.

2.5.3 Branched chain fatty acids synthesis pathway

B. subtilis contains a high percentage of branched chain fatty acids in comparison to other fatty acids and does not depend on the medium condition (Klein et al., 1999). The fatty acid profile of *B. subtilis* is largely dominated by iso- and anteiso-C_{15:0} and C_{17:0} (Toshi Kaneda, 1991). After the production/uptake of branched chain amino acids, it undergoes oxidative desamination and decarboxylation reaction to produce coenzyme A (CoA)-precursors which are used for fatty acid biosynthesis as shown in Figure 18. Based on the amino acids different intermediates are produced, which is finally converted into fatty acids of different types.

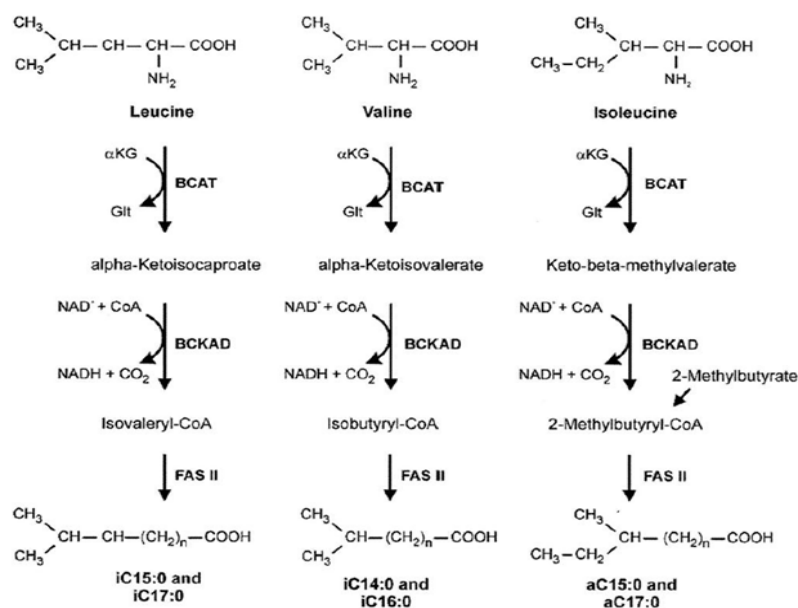


Figure 18 The role of various enzymes in the synthesis of branched chain fatty acids

During fatty acid biosynthesis, valine; leucine and isoleucine are firstly converted to α -Keto-iso-valerate; α -Keto-iso-caproate and α -Keto- β -methyl-valerate, respectively by a soluble branched-chain amino acid amino transferase (BCAT). Then, these intermediates are converted to isobutyryl-CoA; isovaleryl-CoA and 2-methylbutyryl-CoA by a NAD and CoA-dependent branched chain α -keto-acid-decarboxylase (BCKAD). 2-methylbutyryl-CoA is also produced by 2-methylbutyrate. Finally, the activated CoA-precursors are used for the synthesis of various fatty acid species by fatty acid synthetase as shown in Figure 18.

The fatty acid species produced depends on the availability of various substrates that affect the chain isomery and length as shown in the Table 2. In *B. subtilis* branched-chain amino acid amino transferase (BCAT) enzyme uses various α -keto substrates as primers and forms the determining feature of the fatty acid isoforms produced. The reactivity order of the α -keto substrates are as follows α -Keto- β -methyl-valerate \geq α -Keto-iso-caproate \geq α -Keto-iso-valerate (Naik and Kaneda, 1974).

Table 2 Substrates that enhance the synthesis of specific fatty acids in *B. subtilis* (Kaneda, 1977)

SUBSTRATE	EVEN ISO ACIDS (n-C ₁₄ and -C ₁₆)	ODD ISO ACIDS (iso-C ₁₅ and -C ₁₇)	ODD ANTEISO ACID (anteiso-C ₁₅ and -C ₁₇)	EVEN NORMAL- ACIDS (n-C ₁₄ and -C ₁₆)	ODD NORMAL- ACIDS (n-C ₁₅ and -C ₁₇)
Amino acid	L-Valine	L-Leucine	L-Isoleucine	L-Norvaline	L-Norvaline, L- α -aminobutyrate
α-Keto acid	α -Ketoisovalerate	α -Ketoisocaproate	L- α -Keto- β -methylvalerate	α -Ketovalerate	α -Ketobutyrate
Short chain acid	Isobutyrate	Isovalerate	D- α -Methylbutyrate	n-Butyrate	Propionate n-valerate
Others	-	-	L-Threonine α -Ketobutyrate	-	L-Threonine

In *B. subtilis*, there are seven genes present in *bkd* operon that encode proteins which are responsible for the synthesis of branched chain fatty acids through utilization of the respective amino acids. These genes include *ptb*, *bcd*, *buk*, *lpd* (*lpdV*), *bkdA1*, *bkdA2*, and *bkdB*, which encode for phosphate butyryl coenzyme A transferase, leucine dehydrogenase, butyrate kinase, and four components of the branched-chain keto acid dehydrogenase complex: E3 (dihydrolipoamide dehydrogenase), E1 α (dehydrogenase), E1 β (decarboxylase), and E2 (dihydrolipoamide acyltransferase), respectively (Debarbouille et al., 1999). The maximum affinity of this operon is towards isoleucine and valine in comparison to leucine for its activation (Belitsky, 2015).

2.5.4 Branched chain amino acid transporter

Besides the use of branched chain amino acid (BCAA) for the synthesis of their respective fatty acids, these amino acids can also be transported out into the environment. There are various transporters which possess the ability to export as well as import branched chain amino acids from/to the environment (Figure 19). The operon *azlBCDEF* is responsible for both import and export of branched chain amino acids. The gene *azlB* negatively regulates the expression of this operon (Belitsky et al., 1997b). Two genes (*azlC* and *azlD*) of this operon are responsible for the export of branched chain amino acids while *azlE* (*brnQ*) is

responsible for the uptake of valine and isoleucine. The export efficiency of AzlC and AzlD is higher for leucine and isoleucine in comparison to valine (Kennerknecht et al., 2002). The protein, BrnQ depends on Na⁺ for the import of amino acids. Homologues of these proteins were also reported in *C. glutamicum* (Ebbighausen et al., 1989). Although different in functions, the genes *brnQ* and *azlCD* are co-transcribed in *B. subtilis* (Belitsky et al., 1997a).

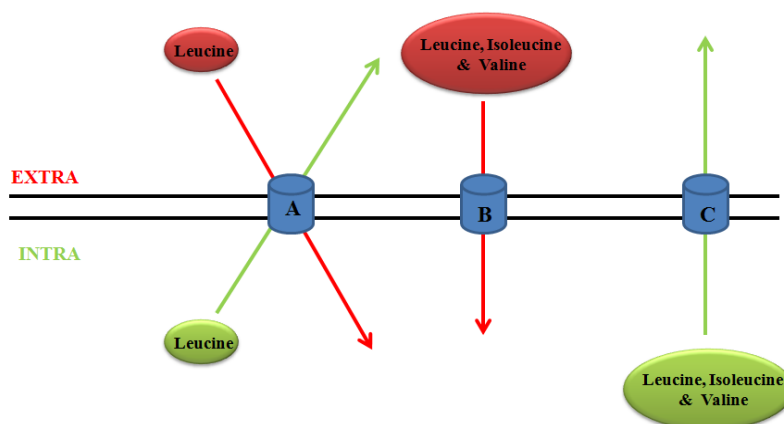


Figure 19 Branched chain amino acid transporters **A)** YvbW (both import and export of leucine); **B)** BcaP, BrnQ and BraB (import of atleast one BCAA) ; **C)** AzlCD export of BCAAs

In addition to BrnQ, BcaP and BraB are also responsible for the uptake of valine, leucine and isoleucine (ILV), while BcaP is the most efficient permease and allows *B. subtilis* to utilize isoleucine and valine as nitrogen sources. In absence of *brnQ*, *bcaP* and *braB*, the activity of CodY was strongly reduced as intracellular concentrations of these amino acids are insufficient to completely activate its regulating activity (Belitsky, 2015). Although the DNA binding activity depends on both GTP and ILV, but prior interaction to ILV is essential for its activity (Belitsky and Sonenshein, 2011). These permeases are adequate enough to provide the cells with ILV from the environment in *liv* mutants of *B. subtilis* which are deprived of ILV biosynthesis when grown in TSS glucose–ammonium medium containing ILV and no growth defects were observed in comparison to the control (Belitsky, 2015). It was reported that the permeases have high affinity towards isoleucine and valine uptake in comparison to leucine, which relates the existence of another

permease which has high affinity towards leucine as a *leuB* mutant strain lacking three permease was able to grow in a minimal medium containing leucine (40 µg/ml) (Belitsky, 2015). A key candidate for this type of permease was found to be YvbW, a putative permease for amino acid which is regulated by T-box mechanism specific to leucine limitation (Rollins, 2014). This permease is responsible for both import and export of leucine.

2.6 Interest of the NRPS peptides produced by *B. subtilis*

Nonribosomal peptides have diverse applications as well as being beneficial for the producer during nutritional starvation and defense mechanism. The amphiphilic lipopeptides i.e. surfactin, fengycin, iturin produced by *B. subtilis* are potential antimicrobial agents, as these compounds have the ability to penetrate and disrupt the cell membrane (Duitman et al., 1999). Toll-like receptors (TLR) are transmembrane proteins, which act as therapeutic agents for the treatment of infectious diseases. Lipopeptides can behave as TLR (Gay and Gangloff, 2007). Lipopeptides have been used to prepare vaccine against hepatitis B virus (Vitiello et al., 1995). The lipopeptide C₁₆-KTTS can be used in skincare application as an antiwrinkle agent (Karl Lintner, 2004). Surfactin, fengycin and mycosubtilin have the potential to be used as biopesticide (Ongena and Jacques, 2008). Besides their function in biocontrol, lipopeptides are involved in bacterial motility and swarming behavior as well as helps in attachment to surfaces (Raaijmakers et al., 2010).

3. SURFACTIN: EXPLORING FROM GENE TO PRODUCT

3.1 Discovery

Arima *et al.* (Arima *et al.*, 1968) isolated an extracellular compound from the growth medium inoculated with *B. subtilis*. Later in the same year, Kakinuma *et al.* (Kakinuma *et al.*, 1968) elucidated the structure of this compound that was named as surfactin. Surfactin is an amphiphilic cyclic lipopeptide consisting of seven amino acids (Glu-Leu-Leu-Val-Asp-Leu-Leu) in its peptide moiety with a chiral conformation of LLDLLDL for the respective amino acids.

3.2 Structure

Surfactin is a branched lactone lipopeptide while the cyclic heptapeptide is linked to β -hydroxy fatty acids with chain length 12-16 carbon atoms. Surfactin has a “horse saddle” topology also known as β -sheet structure. The fatty acid chain can be in *n*, *iso* or *anteiso* configuration while the most predominant chain lengths are C14 and C15. The former is present mainly in *iso*-form or a mixture of *iso/n* while the latter is predominant in *anteiso*-form and *iso*-form is present in fewer amounts (Kowall *et al.*, 1998). The peptide moiety, comprises hydrophobic amino acids at positions 2, 3, 4, 6 and 7 while polar amino acids are at the remaining positions, thus imparting two negative charges to the molecule (Figure 20).

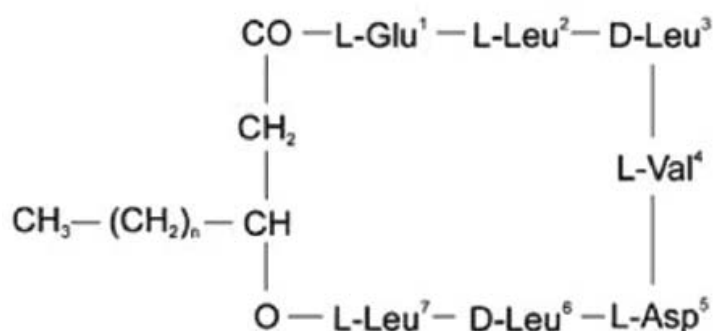


Figure 20 Primary structure of surfactin, where n represents the chain length of fatty acids (Seydlová and Svobodová, 2008)

The three-dimensional structure of surfactin was established using ^1H NMR studies along with molecular imaging. This study revealed the presence of polar residues like glutamic acid (Glu¹) and aspartic acid (Asp⁵) on opposite sides of the molecule. Hydrophobic amino acid (Val⁴) faces the lipidic chain, which forms the major hydrophobic domain as well as Asp⁵ and Leu³ forms the side chain of Val⁴. In addition to L-isoform amino acids, the peptide moiety also includes two residues of D-Leu at position 3 and 6 (Bonmatin et al., 1995). Among all the amino acids in the peptide ring the D-amino acid are highly conserved and plays an important role in structural integrity (Peypoux et al., 1999a). Surfactin has a β -sheet arrangement with a typical horse saddle arrangement, this structural feature accounts for wide range of biological activities (Bonmatin et al., 1994). It has an aggregation number of ~170 and appears as rod-like micelles (Heerklotz and Seelig, 2001).

3.3 Properties

Surfactin are surface-active compounds that can be used as biosurfactants, which are biodegradable and effective in broad range of temperature and pH. It can reduce the surface tension (measure of surface free energy per unit area) of water from $72 \text{ mN}\cdot\text{m}^{-1}$ to $27 \text{ mN}\cdot\text{m}^{-1}$ at a concentration of $20\mu\text{M}$ (Peypoux et al., 1999a). It is the most powerful biosurfactant known till now. Surfactin which is soluble in both polar and non-polar solvents tends to precipitate at lower pH (< 6) as well as in the presence of high salt and organic compounds. In addition to lowering the surface tension, it has an effect on the interfacial rheological behavior by reducing the free energy of the system as well as can influence mass transfer. Microorganisms use this property of the biosurfactant to increase the bioavailability of nutrients through swarming on solid media by attaching and detaching from the substrate. It was observed previously that *B. subtilis* swarming motility is based on flagella and depends on the production of surfactin but *B. subtilis* 168 lacks this motility due to truncated *sfp*, which is essential for surfactin biosynthesis. There are various swarming motility genes i.e. *swrA*, *swrB*, *swrC* (*yerP*) and *efp* whose activation can lead to swarming motility in the laboratory strains. This phenomenon was observed when reversibility of the *swrA* mutation occurred, which enabled swarming motility in *B. subtilis* 168 (Kearns, 2010).

3.4 Biosynthesis

Kluge *et al.* (1988) suggested a multienzyme and, thiotemplate based mechanism known as non-ribosomal peptide synthesis (NRPSs) that catalyzes the biosynthesis of surfactin. A strong regulated operon *srfA* (25kb) is responsible for the synthesis of surfactin (Hamoen *et al.*, 2003).

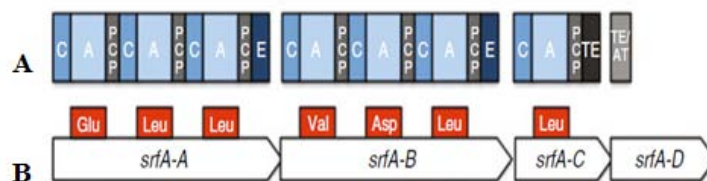


Figure 21 Operon responsible for the biosynthesis of surfactin, **A**: denoting the domains present in each sub-unit, **B**: indicates the amino acid incorporated by each subunit in the final peptide moiety (Jacques,2011)

The polycistronic surfactin operon consists of four modular open reading frames (*srfAA*, *srfAB*, *srfAC* and *srfAD*) (Figure 21). First three open reading frames (ORF) of the operon encodes enzyme SrfAA (402 kDa), SrfAB (401kDa) and SrfAC (144 kDa), these enzymes form the seven modules which consists of unevenly distributed 24 catalytic domains. The seven amino acids constituting the peptide moiety are incorporated by the seven modules whereas the activation, transfer and carboxythioester bond formation is performed by the respective domains of each domain (Seydlová and Svobodová, 2008). The step-wise biosynthesis mechanisms for the synthesis of non-ribosomal peptides have been discussed in the previous section. The protein SrfAA gene is responsible for the activation of Glu, Leu and Leu; SrfAB contains the activation domain for Val and Asp, SrfAC is responsible for incorporation of L-Leu and SrfAD encodes protein, which is structurally similar to thioesterases and belongs to the family of α/β hydrolase (Figure 21).

The initial ORFs (SfAA and SrfAB) are responsible for the elongation through multiple thioester bond cleavage and transpeptidation reactions simultaneously. Besides normal L-isoform amino acids, there are two D-Leu residues. These residues are isomerized by the epimerization (E) domain adjacent to C domain present in both the sub-unit (Peypoux *et al.*, 1999a). The ORF SrfAC is responsible for the condensation reaction

of the last amino acid as well as release of the final lipopeptide intermediate. The final reaction for peptide release is contributed by the thioesterase domain (TE) which is fused to SrfAC by intramolecular lactonization involving a nucleophilic reaction as shown in Figure 22. The macrolactonization process is a stereo- and regiospecific reaction involving the hydroxyl group of the fatty acid and oxygen of 3-(R)-3-hydroxyacyl heptapeptidyl thioester on the Leu⁷ carbonyl group (Bruner et al., 2002).

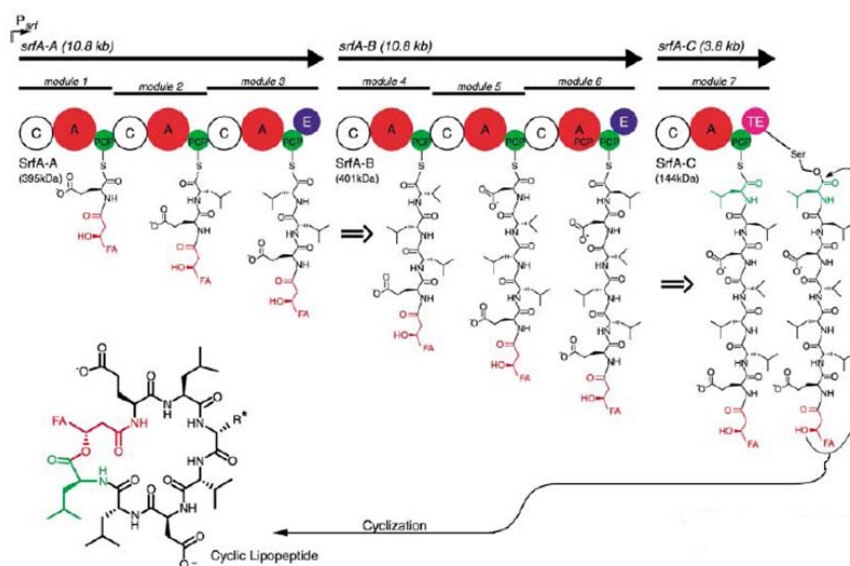


Figure 22 Modular and domain representation for surfactin biosynthesis (Bruner et al., 2002)

The last ORF SrfAD (40 kDa) plays a crucial role of initiation reaction in the biosynthesis mechanism. This protein initiates the biosynthesis mechanism through the formation of β -hydroxyacyl-glutamate by transferring the β -hydroxy-fatty acid substrate to the Glu-module (Steller et al., 2004). This protein also consists of a thioesterase activity (TE), which plays the role of proof-reading and eliminates incorrect transfer of Ppant in the reaction module of the synthetase (Schwarzer et al., 2002). Thus, it can be concluded that synthesis of surfactin requires two bond formations at the extreme ends of the peptide ring: i) carboxyl group of the fatty acid is attached through an amide bond to the amino group of Glu¹; ii) hydroxyl group of the fatty acid is linked through an ester bond to the carboxyl group of Leu⁷.

A comparative study between the activation domain of Surfactin synthetase and Gramicidin S synthetase was carried out. This study revealed that the clusters of amino acid residues responsible for identifying Leu, or Val are similar in all domains as well as Glu-binding region was similar to Asp- and Orn- binding region. It was also found that the Pro-binding domain is also related to Phe-binding domain (Cosmina et al., 1993). This might be the reason for the variability in the surfactin or lipopeptide structure giving rise to various isoforms.

Table 3 Various isoforms of surfactin based on its peptidic moiety along with surfactin (S) native structure

ISOFORM	AMINO ACID TYPE AND POSITION							REFERENCE
	L1	L2	D3	L4	L5	D6	L7	
Surfactin(S)	Glu	Leu	Leu	Val	Asp	Leu	Leu	1
[Ala4]S	Glu	Leu	Leu	Ala	Asp	Leu	Leu	3
[Leu4]S	Glu	Leu	Leu	Leu	Asp	Leu	Leu	4
[Ile4]S	Glu	Leu	Leu	Ile	Asp	Leu	Leu	4
[Val7]S	Glu	Leu	Leu	Val	Asp	Leu	Val	5
[Ile7]S	Glu	Leu	Leu	Val	Asp	Leu	Ile	6
[Ile4,7]S	Glu	Leu	Leu	Ile	Asp	Leu	Ile	7
[Ile2,4,7]S	Glu	Ile	Leu	Ile	Asp	Leu	Ile	7
[Ile2][Val7]S	Glu	Ile	Leu	Val	Asp	Leu	Val	4
[Val2][Ile7]S	Glu	Val	Leu	Val	Asp	Leu	Ile	4
[Val2,7]S	Glu	Val	Leu	Val	Asp	Leu	Val	4
[Ile2,7]S	Glu	Ile	Leu	Val	Asp	Leu	Ile	4

1 (Arima et al.,1968); 2 (Peypoux et al.,1994); 3(Bonmatin et al.,1994); 4 (Peypoux et al .,1991); 5 (Baumgart et al.,1991); 6 (Grangemard et al.,1997)

The adenylation (A) domain present in SrfAA and SrfAB for the activation of L-Glu, L-Leu, L-Asp and L-Leu are highly specific and in vitro studies revealed that the A domain for other positions i.e. 2, 4 and 7 can accept different aliphatic amino acids as substrates (Jacques, 2011). However for the thioesterase domain, it was found that an alteration in the amino acid at positions 2, 3, 4 and 5 are tolerated while position Glu in position 1, and Leu in position 6 and 7 are essential for its acceptance of the substrate (Sieber and Marahiel, 2005). Although a study based on structural content of peptide moiety revealed that position 1, 3, 5 and 6 denoting Glu, Leu, Asp and Leu are usually conservative in nature

and the amino acid present in position 2, 4 and 7 are replaceable (Bonmatin et al., 2003). There can be variation from the native peptidic structure of surfactin giving rise to various isoforms as shown in Table 3 .

3.5 Regulators

The genetic loci mainly responsible for the synthesis of surfactin are *sfp* and *srfA*. Surfactin operon (*srfAA-AD*) expression is regulated by various genes belonging to different phase of microbial growth. The most essential gene regulating its production is *sfp*, this gene encodes an enzyme (224 amino acids) related to the superfamily of 4'-phosphopantetheinases. Both *sfp* and *srfA* are related genetically but are expressed by different genetic loci. It appeared that Sfp regulates the surfactin biosynthetic pathway post-translationally (Cosmina et al., 1993). It acts as a swinging arm between the A and C domain. It is also responsible for the conversion of apo-synthetase to holo-synthetase in *B. subtilis 168* and triggers surfactin production (Lambalot et al., 1996).

During the transition phase, *srfA* expression is induced by two-component system ComP-ComA based on quorum sensing mechanisms triggered by various pheromones. During stationary phase, the expression is regulated by SpoOA-AbrB based on nutritional availability. Another competence related gene *comS*, which is located within *srfA* operon but out of reading frame also regulates its expression (Hamoen et al., 2003). It was found that a null mutant of *spoOH* (encoding σ^H) resulted in a 50% decrease in β -galactosidase activity which relates to *srfA* expression in comparison to *spo*⁺ cells. Thus, it can be inferred that σ^H is responsible for the *srfA* regulation through promoter expression (Nakano et al., 1988). SpoOK has negative influence on *srfA* expression as influx of PhrC is prevented which leads to inhibition of ComA by RapC as this inhibition can be blocked by PhrC (Solomon et al., 1996). Several other transcriptional factors i.e., DegU, PerR and CodY have an influence over *srfA* expression (Serror and Sonenshein, 1996) (Duitman et al., 2007).

According to the Figure 23, bacteria produce various quorum sensing molecules (PhrC, PhrF, PhrG, PhrH and ComX). Accumulation of these molecules triggers a cascade of reaction which regulates the expression of *srfA*. These molecules also help the bacteria to communicate among themselves. Primary regulation is contributed by ComX, which is secreted out of the cell by ComQ. At a particular concentration of ComX, ComP can

autophosphorylate itself as well as phosphorylates ComA. The expression of *srfA* is directly regulated by ComA~P. Besides activating the expression of *srfA*, it also activates ComS, which lies within the surfactin operon but outside the *srfA* ORF. Activated ComS activates ComK, which is essential for competence development as well as release of pheromones. ComS blocks ComK degradation by protease complex ClpCP (Turgay et al., 1997).

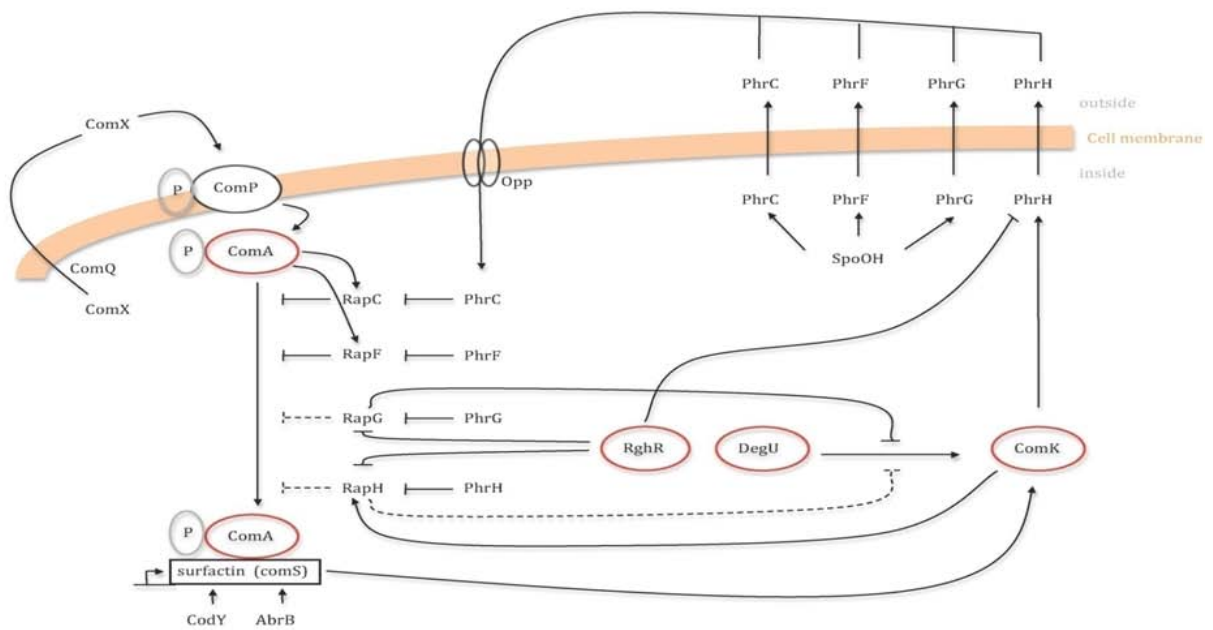


Figure 23 An overview of regulatory pathways affecting surfactin operon expression (Jacques, 2011)

Secondary regulations are contributed by Phr peptides which include PhrC, PhrF, PhrG and PhrH. These small proteins are firstly secreted and upon reaching a certain concentration, Opp (oligopeptide permease) enables the import of these peptides. Phr proteins inhibit the expression of their cognate *rap* genes as *rap* genes hinder the activation of surfactin operon by ComA~P at certain conditions. Although ComA~P can activate the expression of *rapC* and *rapG*.

Surfactin self resistance and efflux is contributed by *yerP*, which has a highest expression at the end of logarithmic phase. Although there is a lack of direct evidence, it was observed that null mutant strain of *yerP* showed growth inhibition at a concentration of 100µg/ml of surfactin as well as surfactin production was decreased in comparison to

control. The expression of this gene only depends on the growth phase and not on the extracellular surfactin concentration. YerP have homology to RND (resistance, nodulation and cell division) family, which can use a proton-motive force efflux pump for the secretion of surfactin. These families of proteins are responsible for oligosaccharides transport, export of heavy metals and solvent tolerance (Tsuge et al., 2001). Recently, a study carried with *B. subtilis* THY-7 a surfactin producer, revealed that *ycxA*, a gene located downstream *urfA* operon might be a putative lipopeptide exporter (Li et al., 2015) .

3.6 Structure and activity relationship

Owing to its amphiphilic nature, it can form dimer with phospholipid bilayer and thus can destabilizing membrane integrity (Carrillo et al., 2003). It can interact with the membrane via hydrophobic interactions, thus affecting the order of hydrocarbon chain as well as membrane thickness (Maget-Dana and Ptak, 1995). Its interaction with phospholipid bilayer is concentration dependent. At low concentration it can penetrate the cell membrane and form mixed micelles with phospholipids; moderate level leads to the formation of ion conducting pores in the membrane through segregation of bilayer; further increase in concentration can result in disruption of membrane displaying its detergent nature (Grau et al., 1999). Below critical micelle concentration (CMC), the lipidic moiety extends freely in solution but above CMC, the lipidic moiety involves in hydrophobic interaction as micelle or oligomers at the interface of two phases (Peypoux et al., 1999a).

The fatty acid branching types have an impact on the surfactant property of surfactin. This property is enhanced in the order *normal* > *iso* > *anteiso*. Based on fatty acid chain length, it was found that increase in chain length leads to an increment in hydrophobicity by a certain factor thus increasing the anti-viral activity. The anti-viral activity differs on the basis of the fatty acid chain length; it was found that C13 fatty acid surfactin isoform showed very less antiviral property in comparison to C14 and C15 isoforms. Monomethyl ester of surfactin C15 was able to inactivate Semliki Forest Virus (SFV) while it was resistant to surfactin isoform mixture (Kracht et al., 1999). However, it was found that the surfactant activity of *n*C₁₄ was higher in comparison to *iso* or *anteiso* C₁₅ (Peypoux et al., 1999a).

The presence of acidic residues within the peptide moiety can assist in membrane penetration. Glutamic acid and Aspartic acid form a claw shaped structure which stabilizes

surfactin-Ca²⁺ complex thus enabling deeper penetrations into the membrane. It can form complex with both monovalent as well as divalent cations, but presence of divalent cations drives complete neutralization (Maget-Dana and Ptak, 1992). When position 2 is occupied by isoleucine in the surfactin variant [Ile², Ile⁴, Ile⁷], It enhances the affinity for calcium in comparison to leucine in position 2 in the variant [Leu², Ile⁴, Ile⁷] as former conformation increases the accessibility of the cation for the acidic residues in comparison to the latter conformation (Kowall et al., 1998). This cation chelation property enables surfactin to inhibit cyclic AMP phosphodiesterase activity (Hosono and Suzuki, 1983). The presence of both isoforms of amino acids (L- and D-) assists surfactin to have specific interaction with its cellular target (Seydlová and Svobodová, 2008). The haemolytic activity of surfactin can be enhanced by esterification of its anionic residues within peptide moiety (Dufour et al., 2005). It was observed that esterification of the glutamate residue resulted in 100% haemolysis at a concentration of 12 µM in comparison to 200 µM of nonesterified surfactin (Kowall et al., 1998). But esterified derivatives show decrease in anti-tumoral activity (Liu et al., 2009). Methylation of the acidic residues enhanced the oil displacement activity by 20% as well as acid tolerance was increased but the anti-viral activity was lost (Shao et al., 2015). Both mono- and di- methylated derivative of surfactin showed decrease and null anti-inflammatory activity respectively (Tang et al., 2010).

The nitrogen source within the nutrient determines the amino acid occurrence in the peptide ring. The position 4 within the peptide ring has most substrate variation. The use of L-leucine or L-isoleucine leads to 5-10% production of [Leu⁴] or [Ile⁴] variant surfactin, whereas the presence of L-valine or L-isoleucine can lead to 40-45% production of [Val⁷] surfactin (Menkhaus et al., 1993). Presence of amino acid in position 4 can increase the hydrophobicity of surfactin in the order Ala < Leu < Val < Ileu which enhanced the surfactant nature of [Ile⁴] surfactin variant. Replacement of hydrophobic amino acid in position 4 (L-Val) can result in a decline in CMC but an increase in monolayer stability at the air/water interface (Bonmatin et al., 1995). Activity comparison based on fatty acid chain length, C14 shows high surface activity whereas the presence of C15 can increase the antiviral and hemolytic activity (Bonmatin et al., 2003). Surfactin isomers with Leu/Ile in position 7 in comparison to Val in the same position have strong inhibitory effect on the release of IL-6 as well as nitric oxide overproduction on LPS-induced murine macrophage cell RAW264.7 (Tang et al., 2010).

The surfactin variant [Δ Asp⁵] has higher haemolytic activity, which led to an enhancement of antimicrobial activity but no zone of inhibition was found for filamentous fungi. Similar studies with [Δ Leu³] and [Δ Leu⁶] resulted in a decline of haemolytic activity in comparison to surfactin but both the variants inhibited filamentous fungi *Fusarium moniliforme* ATCC3893 with MICs of 200 μ g/ml and 50 μ g/ml respectively (Jiang et al., 2016).

3.7 Analysis

Surfactin production can be detected through various methods, which include: i) culture dependent method which comprise haemolytic activity assay, blue agar plate method, atomized oil assay; ii) surface activity methods i.e. cell surface hydrophobicity, surface/interfacial tensions etc iii) on the basis of structure (TLC, HPLC, MS, ECI-MS, MALDI-ToF, NMR) (Plaza et al., 2015). The most commonly used method for separation and analysis of surfactin is RP-HPLC as this method can easily discriminate between various homologs of surfactin with different hydrophilicity. Amino acids present within the peptide moiety based on their hydrophobic properties interact with the mobile phase and the portion with higher hydrophobicity will be eluted at the end i.e. the elution time of surfactin variant should be [Leu7] < [Val7] < [Ileu7]. They can also differentiate based on their fatty-acid chain as the column contains a modified C₁₈ silice, the elution time of the fraction with longer fatty acid chain is longer in comparison to the fraction with shorter fatty acid chain.

3.8 Application

Surfactin has a wide range of application due to its structural arrangement and existence of its monomers in different stereoscopic arrangement as well as presence of fatty acid chain in various lengths. Surfactin has antiviral activity, it can inhibit fibrin clot formation, due to its interaction with lipid bilayer it can form ion channels, it can hinder the activity cyclic adenosine mono phosphate as well as it displays antitumor activity (Seydlová and Svobodová, 2008). Surfactin production helps in pellicle formation for the producer strain at the air/water interface (Chollet-Imbert et al., 2009), enable swarming (Debois et al., 2008) as well as biofilm formation on roots (Bais et al., 2004). Surfactin helps in biofilm

formation which helps in communal development and survival in natural habitat (Stanley and Lazazzera, 2004). The haemolytic behavior of surfactin is the sole drawback characteristics for its medical application. Although it shows haemolytic behavior at concentrations of 40 – 60 μ M (Noudeh et al., 2010), it can completely inhibit mycoplasma growth and proliferation of human colon cancer cells at a concentration of 25 μ M and 30 μ M respectively (Kim et al., 2007). It can also be used for various environmental applications due to its strong surface and emulsification properties. The application includes bioremediation through degradation of hydrocarbons in contaminated sites (Whang et al., 2009); for microbial enhanced oil recovery (Schaller et al., 2004); heavy oil transportation (Ghojavand et al., 2008); xylene water decontamination (Etchegaray et al., 2016), it also stimulates metal remediation by increasing nutrient availability (Mulligan, 2005).

In the following paragraphs, the application of surfactin are categorically explained.

3.8.1 Anti-inflammatory activity

Studies have revealed that surfactin can reduce the impact of inflammatory agent lipopolysaccharide (LPS) on eukaryotic cells. It suppresses LPS activity by inhibiting the expression of IL-1 β and Inos (Hwang et al., 2005) and repress LPS-binding protein (LBP) interaction with lipid A thus hindering the transfer of LPS to its receptor (Takahashi et al., 2006). It was also found that it reduces the levels of nitric oxide, TNF- α and plasma endotoxin in rats due to septic shock (Hwang et al., 2007).

3.8.2 Anti-microbial and Anti-viral activity

Among all the lipopeptides produced by NRPS mechanism, surfactin, iturin and fengycin have wide range of antifungal targets. Surfactin was active against rubberwood sapstain fungus, *Lasiodiplodia theobromae* as it can induce plasomolysis, shrinkage and lysis of fungal mycelium (Sajitha et al., 2016).

Mycoplasma is a causative agent for urogenital tract diseases, respiratory inflammation and also can act as a cofactor in AIDS. Antimycoplasmal activity of surfactin is more desirable in comparison to conventional antibiotics as conventional antibiotics have cytotoxic effects on cells as development of resistant cells. At higher concentrations it can

cause leakage in plasma membrane which leads to influx of medium thus resulting in complete disintegration of the membrane. Surfactin has the ability to inhibit mycoplasma cells with MIC of 25 μ M independent of mycoplasma concentration (Shaligram and Singhal, 2010). Due to low toxicity to mammalian cells, surfactin allows high specificity of inhibition of mycoplasma. Fassi *et al.* showed that when surfactin was used in combination with enrofloxacin it showed two times more activity in comparison to molecules used separately (Fehri et al., 2007).

Surfactin has been found to inhibit biofilm formation for pathogenic organism by preventing them from adhering to solid surfaces or infection sites or on catheters. These pathogenic organisms include *S. typhimurium*, *S. enterica*, *E. coli* and *P. mirabilis*. It was found that surfactin at concentrations between 5-50 mg/l can inhibit biofilm formation of *Salmonella* on catheter (Mireles et al., 2001). It can also restrict growth of *P. syringae* on roots (Bais et al., 2004). Surfactin surface-active property enabled the producing organism to spread over wood blocks after [one+](#) week of inoculation even for inhibiting the growth of *Lasiodiplodia theobromae* (Sajitha et al., 2016).

Anti-viral activity of surfactin includes inactivation of membrane integrity by incorporating into lipid bilayer leading to disintegration of the envelope and it was found that surfactin was more efficient against enveloped viruses than non-enveloped viruses. The reason behind such activity is the fact that antiviral activity depends on viral envelope lipid composition as well as glycoproteins encoded by the virus. The activity can be against wide range of viruses i.e. Semliki Forest Virus, herpes simplex virus, feline calicivirus, murine encephalomyocarditis virus, etc. (Vollenbroich et al., 1997). Anti-viral study was carried out on three viruses: vesicular stomatitis virus (VSV), suid herpes virus type 1 (SHV) and Semliki Forest Virus (SFV). It was found that C13-surfactin has very less anti-viral activity only against SHV while C14- and C15-surfactin were highly active against both SHV and VSV. But all the three isoforms were inactive against SFV. When the isoforms were esterified and added in equimolar amount, it was found that C14 and C15 monomethyl ester were active against both SHV and VSV. While in the case of Semliki Forest Virus, it was observed that the monomethyl C14 and C15 showed anti-viral activity but no activity was observed for non-esterified surfactin mixture (Kracht et al., 1999). Thus, it can be concluded that the anti-viral activity depends on fatty acid chain length (hydrophobicity), peptide moiety charge as well as on virus species.

3.8.3 Biomedicine

It was reported that surfactin can act as antitumor agent against Ehrlich's ascite carcinoma cells (Kameda et al., 1974). It can inhibit human colon carcinoma cell proliferation by inducing apoptosis as well as by suppressing the signals of cell cycle (Kim et al., 2007).

It can inhibit fibrin clot formation both *in vitro* and *in vivo* by inhibiting the conversion of fibrin monomer to its respective polymer (Arima et al., 1968). It increases the activity of prothrombinase as well as changes the conformation of plasminogen at a concentration of 3-20 μM . It can also avoid platelet aggregation (Kikuchi and Hasumi, 2002). This activity of surfactin is due to its effect on the downstream signaling pathways (Kim et al., 2006). Surfactin display a notable antihypercholesterolemic activity (Arima, *et al.*, 1972).

3.8.4 Biocontrol

Around 50% of the biopesticides in world market are produced industrially by species of *Bacillus* (Fravel, 2005). They oppose pathogens for niche and nutrients as well as can activate defensive mechanism of the host plant and direct them to produce antimicrobial compounds. *B. subtilis* can produce wide variety of bioactive compounds and surfactin is among one of those compounds. These compounds are highly promising in comparison to chemical counterparts for agricultural application because of their high biodegradability, low toxicity and eco-friendly properties. These compounds can hinder the growth of phytopathogens and pests causing plant diseases (Ongena and Jacques, 2008).

Surfactin's role as biocontrol can be subdivided into two categories; firstly, its role to stimulate systemic resistance in plants (Jourdan et al., 2009) and secondly, as a factor to enhance colonizing efficacy as plant growth-promoting rhizobacteria (PGPR) (Gao et al., 2016).

Basal immune response in plants is triggered upon recognition of microbial motifs known as pathogen-associated molecular patterns (PAMPs). These motifs are generally recognized by high affinity pattern recognition receptors (PRR) located in the plasma membrane of plants (Henry et al., 2011). These receptors consist of extracellular ligand-binding domain, single transmembrane domain and an intracellular kinase-signaling domain. Induced systemic resistance is triggered by a local stimulus by a beneficial isolate/compound which results in extension of resistance in all the organs of the plant.

Surfactin can trigger defense-related responses in tobacco suspension through ROS accumulation both in the cytoplasm and extracellular environment. ROS formation is induced through plasma-membrane associated NADPH oxidase quenched either by cell wall or extracellular peroxydases. At a concentration of 20 μ M of surfactin, no cell death was observed but can develop signaling process which can activate biochemical pathways leading to development of defensive response (Henry et al., 2011). Surfactin can turn on oxylipin pathway in plants, which can produce various biologically active secondary metabolites resulting from unsaturated fatty acids. These metabolites may accumulate at the infection site and inhibit pathogen entry. Surfactin can induce the production of other secondary metabolites which resemble to phenylpropanoids deriving from PAL enzyme. The plasma membrane requires specific composition and/arrangement of the lipid bilayer for the perception of surfactin. Surfactin is recognized as PAMPs at the cell surface and not as effector molecule transported into the cell. However in comparison to PAMPs which are recognized in nanomolar concentration, surfactin is needed in micromolar range (Jourdan et al., 2009).

B. subtilis can stimulate plant growth and control plant diseases through root colonization. Root colonization is facilitated through a swarming mechanism. Swarming mechanism in *B. subtilis* is directed by *swrA* operon (Kearns et al., 2004) and flagellum as well as surfactin is also required for this phenomenon. Surfactin promote swarming by reducing the surface tension between the bacterial cell and substrate as well as it favors grouping of cells into dendrites (Kearns, 2010). It was reported that surfactin is important for biofilm formation on *Arabidopsis* root surfaces by *B. subtilis* 6051 (Bais et al., 2004). Impairment of surfactin synthesis in mutant Δ *srfAC* can obstruct swarming mechanism. which can lead to reduced root colonization efficacy (Gao et al., 2016). This suggest that profitable outcome of PGPR in agricultural use can be increased by boosting the swarming motility.

4. OVERPRODUCTION OF SURFACTIN

4.1 Present scenario

The surfactant industry is increasing by 9 million US\$ every year. The surfactants mostly available in market are of petrochemical origin. The previous chapter concluded with the broad range of applications of surfactin (a biosurfactant) due to which it is gaining importance over chemical surfactants. A study was carried out to synthesize this molecule through a chemical process. Although it was revealed that the synthetic process requires multiple steps and ultimately increasing the production costs. Despite surfactin's advantage over its chemical origin counterpart, the use of surfactin for commercial application is very limited due to its low-level of production in comparison to high-level requirement in specific applications. Sigma-Aldrich markets surfactin at a price of 25.4€/mg for research purpose. Thus, the problem needs a biotechnological solution to make production cost economical and sustainable. Lipopeptide production in *B. subtilis* is closely linked to competence and sporulation, thus forming a three way metabolic network (section 3.5). It was inferred from various studies that surfactin production depends on rate of utilization of carbon source, oxygen availability and factors governing lipid formation (Peypoux et al., 1999b).

4.2 Role of various factors

Various strategies like optimizing the medium, using waste from agriculture or industrial origin to reduce raw material cost or bioreactor design have been utilized to reduce the production cost. Implementing various mutagenesis techniques for strain improvement has also been carried out to overproduce surfactin (Kim et al., 1997).

4.2.1 Nutrient composition

Various culture media are used for the production of surfactin i.e. Cooper's medium, Landy medium etc. The production of surfactin starts from exponential phase and maximum production is observed during transition phase. It has already been seen in the previous chapter (Chapter 1.3 - Regulators) that surfactin production depends on cell biomass and biomass requires proper nutrient composition for its growth. Thus, nutrient

composition can have a profound effect on the surfactin production. *B. subtilis* ATCC 6633 can produce surfactin when grown on Landy medium (Duitman et al., 1999). The dependence of nutrient composition is not only limited to carbon sources but also various nitrogen sources and various minerals have a significant impact on the production. An alteration of medium composition for enhancing surfactin production is an effective approach.

Carbon source: Glucose is the most preferred carbon source for surfactin production. Mainly, glucose is used as carbon source with a concentration between 20 and 40 g/l. *B. subtilis* C9 when grown in the presence of glucose, the specific surfactin yield was higher in comparison to sucrose as carbon source (Kim et al., 1997). Various complex carbon sources i.e. diesel, starch, glycerol or feed stock have also been used for the study of surfactin production (Etchegaray et al., 2016) (Amin, 2014). In the absence of carbon source surfactin production was not observed as the cells were not able to grow properly. When cells were grown with an excess of glucose (80 g/l), the surfactin yield was significantly decreased, which can suggest that an excess of glucose can have a negative impact of the production (Dauner et al., 2001). Utilization of glucose as a sole carbon source can lead to a production of 80 mg/l of C₁₅-surfactin from *B.amyloliquefaciens* MB199 (Liu et al., 2012).

Nitrogen source: Cooper medium contains ammonium nitrate and classical Landy medium contains glutamic acid as nitrogen source. Both medium have been widely used for surfactin production. Thus, it can be revealed that there is no preference for nitrogen source for surfactin production. *B. subtilis* ATCC 21322 utilizes organic nitrogen for growth, while the inorganic source i.e. nitrate is utilized for switching on secondary metabolism. In aerobic condition, nitrate utilization can only occur when ammonium is depleted (Davis et al., 1999). When a derivative strain of *B. subtilis* 168 was grown in Landy medium for 24 hours in Biolector[®] device (48-well microtiter Flowerplates), it was observed that higher concentrations of glutamic acid can have a positive impact on surfactin production (Motta Dos Santos et al., 2016). When the nitrogen source of the cooper medium was substituted from NH₄Cl to (NH₄)₂SO₄, an enhancement of surfactin production was observed (Willenbacher et al., 2015). Addition of (NH₄)₂SO₄ enhanced the C₁₅-surfactin production from *B.amyloliquefaciens* MB199 upto 60 mg/l (Liu et al., 2012). Wie et al (2004) found that iron and manganese can boost the production of surfactin. It was determined that a critical molar ratio of 920:7.7:1 for

nitrogen/iron/manganese is required for continuous production (Sheppard and Cooper, 1991).

Other nutrients: These nutrients include magnesium, manganese, potassium, iron etc. Higher concentration of both Fe^{2+} and Mn^{2+} can stimulate surfactin production but the former can produce phthalic acid which can decrease the pH and surfactin solubility (Wei et al., 2003), while the latter requires cooperation from nitrogen source and other essential factors to have a positive effect (Wei et al., 2007). However addition of 4 mM of Fe^{2+} along with pH adjustment can enhance the surfactin production to 3 g/l in minimal salt medium with 40 g/l glucose using *B. subtilis* ATCC 21322 (Wei et al., 2004). When the same strain was cultivated in the presence of Mn^{2+} concentration (0.001 to 0.1 mmol/l), the maximum production of 1.5 g/l of surfactin was obtained with 40 g/l of glucose. Upon increasing the Mn^{2+} concentration to higher than 0.05 mmol/l, a shift of nitrogen utilization was observed as nitrate became the main nitrogen source instead of ammonium. Addition of Mn^{2+} increases both glucose usage and nitrate reductase activity. The increase in nitrate reductase activity supports the nitrate utilization efficacy of the cells, thus leading to highest surfactin production during nitrate-limited conditions. Glutamate synthase (GOGAT) which is responsible for the synthesis of glutamate, the activity of this enzyme was enhanced in the presence of manganese ions. This process led to the generation of free amino acids for surfactin synthesis (Huang et al., 2015). In the presence of glucose (40 g/l), both Mg^{2+} and K^+ ions can have significant impact on *B. subtilis* ATCC 21322 for surfactin production. Upon decreasing the magnesium concentration from 0.6 mM to 0.04 mM, the surfactin concentration decreased from 0.69 g/l to 0.16 g/l. When K^+ concentration was increased from 0 to 10 mM, both increase in cell growth and surfactin concentration were observed. Maximum surfactin production of 3.34 g/l was obtained with potassium concentration of 10 mM. According to Taguchi and statistical analysis, it was found that Mg^{2+} and K^+ ions play a greater role in surfactin production in comparison to Mn^{2+} and Fe^{2+} . The impact of Mg^{2+} and K^+ can affect the cell as the former is a cofactor of Sfp protein which is essential for surfactin production, while the latter helps in surfactin secretion (Wei et al., 2007).

4.2.2 Oxygenation

Dissolved oxygen and mechanical agitation plays a determining role for surfactin production. Surfactin production is highly linked to oxygen volumetric mass transfer coefficient (K_{La}) under various agitation and aeration rate. The highest surfactin production (7g/l) was obtained in the foam using *B. subtilis* C9 in the presence of Cooper's nitrogen source and oxygen limiting conditions (Kim et al., 1997) but *B. subtilis* S499 requires high aeration condition (Jacques, 2011). Maximum surfactin production was obtained at a K_{La} value 47.52 h^{-1} . Since surfactin is a growth associated product, the production rate is directly proportional to the specific growth rate of the cell. So, aerobic bacteria like *B. subtilis* needs adequate oxygen concentration for growth and product i.e. surfactin formation. At a fixed aeration rate, K_{La} can be increased on increasing the agitation rate in bioreactor or flask culture. At a low aeration rate of 0.5 vvm, maximum surfactin production increased 1.5 folds on increasing the agitation rate from 200 rpm to 350 rpm during batch fermentation (Yeh et al., 2006). Thus, in a batch mode both aeration and agitation rates have a great impact on cell growth and surfactin production.

The volumetric oxygen transfer co-efficient in Biolector with each filling ratio was calculated with the following equation,

$$K_{La} = 6.67 * 10^{-6} N^{1.16} V_L^{-0.83} d_o^{0.38} d^{1.92} \dots\dots\dots 1$$

Where N is the agitation speed (min^{-1}); V_L is the filling volume ratio; d_o is shaking diameter; d is the maximum inner diameter of flask (Fahim et al., 2012a). The agitation speed, shaking diameter and inner diameter of the flask was kept constant during the entire study with values 250 rpm, 5 cm and 12 cm, respectively.

The volumetric oxygen transfer co-efficient of the Biolector was calculated according to the equation,

$$K_{La} = OTR_{\max} / L_{O_2} * p_G \dots\dots\dots 2$$

Where OTR_{\max} is maximum oxygen transfer capacity; oxygen solubility ($L_{O_2} = 9.89 * 10^{-4} \text{ mol/L/bar}$ at 37°C) and partial pressure of oxygen in gas phase ($p_G = 0.2095 \text{ bar}$) (Funke et al., 2009)(Gnaiger, 2001). The working volume used in Biolector was 1ml with total well volume 3.6 ml.

High shaking rate in flask condition is beneficial for the production of lipopeptides by *B. subtilis* (Jacques, 2011). As in shaking flasks, the overall oxygen absorption rate depends on the replacement of thin liquid layer or fresh liquid surface generation. It was found that poor oxygenation has a detrimental effect on surfactin production while K_{La} from 10.8 to 288 h^{-1} have beneficial impact on surfactin production. It was observed that a $K_{La} > 216 \text{ h}^{-1}$ can significantly enhance the surfactin production of a derivative strain of *B. subtilis* ATCC 21322 (Fahim et al., 2012a). Based on filling volume, it was found that when the filling volume was increased there was a significant decrease in the surfactin production by *B. subtilis* ATCC6633 while the filling volume did not show significant impact on the production by a derivative strain of *B. subtilis* ATCC6633, where the native promoter was replaced by a constitutive one (Guez et al., 2008). Lipopeptides especially surfactin can decrease K_{La} owing to its high surface activity as it forms a layer of stable film in the interfacial area prohibiting oxygen transfer into the medium (Hbid et al., 1996). It can be hypothesized that if the inner diameter of the flask can be increased this surface activity of surfactin can be reduced, which can lead to higher K_{La} and thus increase the surfactin production.

4.2.3 Strain improvement

A surfactin overproducing mutant strain from a wild type *B. subtilis* (B.s-E-8) strain was obtained using ion beam implantation. This strain was cultivated in a modified bioreactor with cell/foam recycler. The surfactin production reached to a maximum value of 12.20 g/l after 32h cultivation (Gong et al., 2009). An improvement of 4-6 times and 4-25 times was obtained using random mutagenesis with *N-methyl-N'-nitro-N-nitrosoguanidine* on *B. subtilis* ATCC 55033 and *B. subtilis* SD901, strain respectively. Implementing UV mutagenesis generated an overproducer strain of *B. subtilis* ATCC 21332. This mutant known as ATCC 5138 has 30 times low isocitrate dehydrogenase activity in comparison to its parent strain and can produce 1.1 g/l of surfactin (Carrera and Cosmina, 1993).

In the following sections, various genetic engineering approaches for strain improvement for overproduction of surfactin have been discussed through direct and indirect approaches.

4.3 Direct approach

Direct approach outlines distinct strategies used for strain improvement through modification of the native *surfA* gene or replacement of the native promoter and lastly a hypothesis has been developed on the basis of *sfp*.

4.3.1 Role of promoters

The promoter forms the most important component in an expression system as it can dictate the expression level of a gene. The activity of a promoter depends on the consensus sequence in the several core regions known as UP element. This UP element consists of -35 region, -15 region, -10 region and +1 region (Cheng et al., 2016). Till date, there are three categories of promoters used for the overproduction of a particular molecule. This includes autoinducible promoter, constitutive promoter and inducer specific promoter.

Autoinducible promoters can trigger expression of target gene based on the growth phase (late log phase to stationary phase) in the absence of an inducer. The promoters under this category includes P_{pst} , P_{rpsF} and P_{ppsA} (Cheng et al., 2016). Constitutive promoters are those promoters which can activate expression of a gene irrespective of growth phase or any stimulus. P_{veg} (Lam et al., 1998), P_{repU} (Coutte et al., 2010), P43 (Ye et al., 1999), P_{secA} (Herbort et al., 1999), P_{yxiE} (Zhang et al., 2007) etc. comes under this group of promoters. Lastly, the inducer specific promoters require inducer for the expression of gene under their control. This promoter is widely used for the production of various biological molecules but is not cost effective. The examples of this promoter are *spac* promoter, T7 promoter, *groES-groEL* promoter, driven by IPTG; *xylA* induced by xylose (Cheng et al., 2016). Surfactin production of 50 g/l was obtained with *B. subtilis* strain having exogenous *lpa-14* gene. The culture was carried out in a fermentation vessel and the surfactin was analyzed in the liquid medium without the removal of foam (Tadashi Yoneda et al., 2006). Replacement of native promoter with a constitutive promoter, P_{repU} (Coutte et al., 2010) and with an inducible promoter, P_{spac} has been studied (Sun et al., 2009).

The role of each type of promoters for the overproduction of surfactin have been described precisely in the following paragraphs.

The promoter for surfactin operon P_{srfA} is an autoinducible expression system activated by signal peptides i.e. ComX and CSF (Figure 23). These signal molecules are components of quorum sensing pathway leading to competence development. The promoter's activity depending on cell density was studied using GFP fluorescence, the expression was found maximum during the transition phase, constant during stationary phase and low during early and mid-exponential phase (Guan et al., 2015).

The promoter consists of 607bp and regulatory regions have already been determined located at -35 region and -10 region, which is the transcriptional start site (Nakano et al., 1991). Its 5'-flanking region spans from -607 to -40 bp. The expression of various truncated region of the 5'-flanking region was studied in fusion with GFP. This study revealed that the 5'-flanking region (-136 to -76) showed significant reduction in the GFP fluorescence intensity, which determines its significance for the expression of the promoter. The maximum expression was obtained with the mutant having the region (-136 - +291) in comparison to wild type and this promoter with this sequence (AATGTT.....A.....ATGACAATG) was termed as P04. Site-directed mutagenesis was carried out in the region between -35 and -10 of the promoter P04 (GTGATAAAAACATTTTTTTCATTTAAACT); this generated a promoter P10 (TTGACA) with a change in the consensus sequence at -35. The expression P10 was maximum among all derivatives of P04 as well as the expression was 1.5 times more than P04 (Cheng et al., 2016).

Besides modification of the UP element, various studies have been performed where at least two promoters (same or different) were placed as tandem repetitive sequences upstream of the expressed gene. The productivity of TSaGT in *B. subtilis* was enhanced several folds by the alignment of *HpaII* promoter along with *blma* promoter or *amyR2* promoter (Kang et al., 2010). Alignment of same promoter was studied by aligning MCP*tacs* promoter clusters until it reached the critical number of five (Li et al., 2012). This strategy enhances the window of expression of the system. Guan et al. (2016) used two way strategies to extend the window of expression *srfA* operon in a non-sporulating *B. subtilis* strain, where the σ^F was deleted. First strategy was to link *gsiB* promoter to P_{srfA} , where the former is under σ^B and the latter is under σ^A . This strategy did not enhance the activity in comparison to the control. Whereas the other strategy, P_{HpaII} to P_{srfA} , where both the promoters are under the control of σ^A enhanced the activity to a significant amount. The consensus sequence was also changed in both promoters resulting in higher expression

activity in comparison to respective control (Guan et al., 2016). Thus, it can be concluded that changing the core regions of the promoter can lead to overexpression of surfactin as well as it can help in tuning the synthetic genetic circuits. While the use of dual promoters can increase the expression window time of *srfA* operon.

In this section the role of constitutive promoter for the production of surfactin has been discussed. Coutte *et al.* (2010) constructed a mutant strain in which the native promoter was replaced with constitutive promoter P_{repU} (BBG 113), a promoter which regulates the replication gene *repU* in *S. aureus* as well as a mutant strain with same promoter construction but the *ppsA* gene was disrupted (BBG 131). Both mutant strains were obtained in *B. subtilis* 168 background and the control strain was named as BBG 111. To restore lipopeptide production, *sfp* gene from *B. subtilis* ATCC21332 was inserted in order to convert the Apo-synthetase to Holo-synthetase. The promoter substitution increased the surfactin production five times during the early stage of growth. Continuing the incubation period, it was found an enhancement of 1.3 times of surfactin production for BBG 113, and an increase of 1.2 times with BBG 131 in comparison to BBG 111 at 24 hours in Landy medium. Consequently the haemolytic activity was enhanced; the maximum halo diameter was obtained with BBG 131 after 3 days of incubation. Another constitutive promoter P_{veg} which is endogenous promoter of *B. subtilis* regulating vegetative gene *veg* (Radeck et al., 2013). This promoter was integrated in the upstream region of *srfA* operon in two strains *B. subtilis* 3A38 and *B. subtilis* DSM 10^T. The former originating from strain NCIB 3610 produces surfactin in small amount while the latter is a wild type, high surfactin producing strain. The mutant strain of 3A38 produces surfactin to a maximum concentration of 0.26 g/l (3.7 times higher than control) while the mutant of DSM 10^T could only produce a mere quantity of 0.04 g/l (15.5 times less than control). It was concluded that the surfactin biosynthesis regulation is different in dissimilar strain based on its ability to produce surfactin without modification. Introduction of the constitutive promoter in the high producer DSM 10^T led to a disturbance in quorum sensing mechanism which ultimately resulted a decline in surfactin production (Willenbacher et al., 2016a).

Lastly in this segment, the role of inducible promoter for overproduction of surfactin has been discussed. Replacement of the native promoter by an inducible promoter was carried out in *B. subtilis* *fmbR* to generate a mutant strain *fmbR*-1. It was found that *fmbR*-1 was able to produce 1.56 g/l of surfactin without induction while with an induction of 300

$\mu\text{mol/l}$ of IPTG, it was able to produce 3.86 g/l and with lowest induction of $100\mu\text{mol/l}$ of IPTG *fmbR-1* was able to produce 1.39 g/l of surfactin. In comparison to control, without induction, *fmbR-1* produces 4 times higher while with induction 10-fold more with $300\mu\text{mol/l}$ of IPTG. Beyond $300\mu\text{mol/l}$ of IPTG induction, the surfactin production was decreased as high induction might lead to spore formation before surfactin production. The mutant strain's antagonistic activity against *B.cereus* was also increased (Sun et al., 2009).

4.3.2 Role of *Sfp*

Phosphopantetheine-transferase plays a crucial role in lipopeptide production through NRPS mechanism. *Sfp* encoded by genetic loci *sfp*, transfers 4'-phosphopantetheine moiety from coenzyme A (CoA) onto the conserved serine residue of peptidyl carrier protein (PCP). In *B. subtilis*, there are several carrier proteins which include: acyl carrier protein (ACPs) for fatty acid synthesis, ACPs for polyketide synthesis, PCPs for nonribosomal peptide synthesis (Walsh et al., 1997). A comparative study was carried out between acyl carrier protein synthetase (AcpS) and *Sfp* on the basis of their specificity to transfer CoA to various protein partners ACP, DCP and PCP. AcpS was able to modify only ACP and DCP required for primary metabolism and not PCP which is required for secondary metabolism. On a contrary, *Sfp* recognizes all proteins and enables its modification (Mootz et al., 2001).

Overproduction of *sfp* gene has a negative effect on the expression of *srfAA*. This was reflected in a reduction of surfactin production too. Thus it can be inferred that *sfp* can regulate and interact directly in surfactin synthesis (Nakano et al., 1992).

Thus, it can be hypothesized that modification of *sfp* or binding site of cofactor on ACP and DCP restricts its substrate specificity to PCP, and it might lead to higher Phosphopantetheine-transferase in NRPS, which results in overproduction of surfactin. Modification of *sfp* promoter is required as it was found out that it is under the control of a weak constitutive promoter (Nakano et al., 1992).

4.3.3 Other factors

Bacillus subtilis 168 can produce two families of lipopeptides namely surfactin and plipastatin (fengycin) (at very low concentration, under 50 mg/l) when *sfp* gene is

integrated into its chromosome as *B. subtilis* harbors a mutated *sfp* gene. In order to direct the precursor flux towards the production of one lipopeptide, a study was carried out with disruption of *ppsA* gene which is responsible for plipastatin synthesis. The mutant strain produced lipopeptides 1.8 times in Landy medium and 1.03 times in Landy MOPS medium in comparison to control at 24 hours (Coutte et al., 2010).

Replacement of native A-PCP domains with A-PCP domains of different bacteria and fungus origin recognizing different amino acids lead to the production of different surfactin isoforms with variant properties, although there was a marked reduction in the production of surfactin due to the specificity of the C-domain acceptor site (Peypoux et al., 1999a).

4.4 Linkage between Direct and Indirect approaches

The expression of *srfA* operon is activated by a two component system ComA/ ComP and inhibited by CodY and AbrB through their binding to the promoter of *srfA* (Hamoen et al., 2003). Through genome shuffling technology, an overproducing *B. amyloliquefaciens* mutant FMB38 strain was developed from parental strain ES-2-4 (Zhao et al., 2014). The production of surfactin in the mutant strain was increased 3.5 times and 10.3 times in shaken flask and fermenter respectively. Expression analysis revealed that the *srfA* transcript number of mutant strain was 15.7 times higher than the parental strain. Through comparative proteomics methodology differentially expressed proteins were analyzed. Three proteins ComA, DegU and CodY directly related to surfactin synthesis were differentially expressed. It was found that the ComA was unregulated which was obvious as ComA binds to *srfAD* promoting its expression and thus increasing surfactin production. Downregulation of DegU was observed. But a contradictory result was observed with the upregulation of CodY as it was found earlier that CodY acts as a transcriptional inhibitor of *srfA* expression (Serror and Sonenshein, 1996b) but *codY*-knockout strain of *B. subtilis* 6633 showed no effect on surfactin production (Duitman et al., 2007). In order to enhance the surfactin production it was observed that not only enhancement of promoter activity but also overexpression of various proteins is required. Thus proving the role of indirect approach can also lead to surfactin overproduction which includes surfactin monomer overproduced which has been discussed in the following section of the chapter.

4.5 Indirect approach

This approach deals with various aspects of overproduction of surfactin based on overproduction of its monomers. *B. subtilis* can utilize diverse carbon sources. Lipopeptide synthesis was carried out *in vitro* in the presence of amino acids, ATP, CoA-iso-3-hydroxytetradecanoic acid (Ullrich et al., 1991). The expression of *urfA* depends upon the phosphorylation of ComA by ComP. The genetic loci *pta* is responsible for the conversion of acetyl-CoA to acetyl phosphate as well the reverse reaction of *ackA* can also produce acetyl phosphate from acetate. This acetyl phosphate acts as a phosphate donor for the phosphorylation of ComA (Kim et al., 2001). As a result, *pta* and *ackA* can indirectly activate the expression of *urfA*. In a surfactin overproducing strain, there was an upregulation of 2-Aminoethylphosphonate-pyruvate transaminase, which is involved in the synthesis of aspartate (Zhao et al., 2014). The peptide moiety of surfactin contains aspartate, so an enhancement of aspartate can result in an increase of surfactin production. So we can speculate that an abundance of key enzymes or specific enzymes can have a positive impact on surfactin production.

Liu *et al.* (2012) carried out a feeding experiment with different amino acids supplied to the media and the surfactin variants were analyzed using *B. subtilis* TD7 strain. Addition of amino acids not only affects the amino acids moiety in the peptide ring but also affects the hydroxyl fatty acid moiety. It was found that addition of L-valine to the culture can lead to the production of [Val⁷] variant surfactin (Peypoux et al., 1999c). Leucine and valine are the precursors for odd and even *iso*- β amino acid and fatty acid, respectively while isoleucine is the precursor for odd anteiso-amino acid and fatty acid (Hourdou et al., 1989). When no amino acid was added to the culture medium the proportion of surfactin variants was found to be C15>C14>C13>C16 based on β -hydroxy fatty acid isoform. Feeding with valine (1 g/l) maximum proportion of *iso*-C14 was observed in comparison to other surfactin variants. In the case of Leucine, maximum proportion of *iso*-C15 was observed in comparison to others but upon addition of Isoleucine the maximum proportion was found to be *anteiso*-C15 (Liu et al., 2012).

Thus, it can be concluded from chapter 1 as follows:

1. *B. subtilis* can be represented as a model for gram positive bacteria and is highly efficient for the production of high valued products.

2. NRPS is a RNA- independent mechanism which depends on various monomers for the assembling of various non-ribosomal peptides.
3. Surfactin is a highly-valuable end product produced by *B. subtilis* through NRPS mechanism with highly complex regulatory pathway.
4. Various approaches have been carried out to overproduce surfactin, but until now no evidence have been provided for surfactin overproduction based on intracellular increase of its precursors, as for example leucine, which is one of the monomers required for non-ribosomal peptide synthesis of the peptide moiety of this lipopeptide.

Here, a study was carried out based on the hypothesis that the intracellular pool of branched chain amino acids has determining effect on the production of surfactin.

CHAPTER 2: MATERIALS AND METHODS

MATERIALS

1 Strains

1.1 *Bacillus subtilis* ATCC 168 and *Bacillus subtilis* BSB1

In this entire study, two wildtype strains of *B. subtilis* (*BS 168* and *BSB1*) were used for the construction of various mutants and for the analysis of surfactin production. The strain *BSB1* is a tryptophan prototrophic (tryp^+) obtained from *BS168* strain, which is tryptophan auxotroph (tryp^-). *B. subtilis* BSB1 strain (BaSysBio reference strain) was used for the transcriptomic study exposing it to various environmental and nutritional factors that it can confront in the living world (Nicolas et al., 2012). It can be deduced that both the strains have genomic similarity. The entire genome of *BS 168* has already been sequenced and is easily accessible with accession no. NC_000964.3 (NCBI database). It possesses high specificity and efficiency of genetic transformation and is widely used as a genetic model. The genomic analysis revealed, that contains two operons encoding for nonribosomal peptide synthetases: the surfactin and plipastatin operons. But it cannot produce the lipopeptides as it has mutated *sfp* gene, which is responsible for the conversion of apo-synthetase to holo-synthetase). The *sfp* gene encodes a 4-phosphopantetheinyl transferase, which converts inactive apoenzyme peptide synthetases to their active holoenzyme forms by posttranslational transfer of the 4-phosphopantetheinyl moiety of coenzyme A to the synthetases. An intact *sfp* gene (844 bps) was transferred from *B. subtilis* ATCC 21322 to BS168 and BSB1 strains to convert them into lipopeptide producing strains BBG111 and BBG258 respectively (Coutte et al., 2010).

1.2 *Bacillus subtilis* BSB1N

Strain improvement through metabolic engineering is thought to be a well suited strategy for the overproduction of a particular metabolite. In order to carry out metabolic engineering various deletions and insertions of genes need to be carried out for appropriate flow of metabolite in a desired direction. Availability of selective markers forms one of the crucial limiting factors to carry out multiple modifications. In order to carry out multiple modifications, markerless strategies are required. There are various strategies which carry out modification in a markerless fashion i.e. temperature sensitive plasmid (pSC101

replicon) for gene replacement (Hamilton et al., 1989); gene replacement using pORI vectors (Leenhouts et al., 1996); gene replacement using auxotrophy based method (Brans et al., 2004); Cre/lox system for genome engineering (Yan et al., 2008).

Finally, a strategy based on Pop-in Pop-out technique using alternative neomycin and phleomycin selective marker (Tanaka et al., 2013) was found impressive to carry out the modification. In order to carry out this technique, a strain TF8A lacking 231 kb of chromosome in comparison to the wildtype *B. subtilis* 168, which contains the prophage SP β and PBSX as well as prophage-like element skin (Westers et al., 2003). The *upp* gene of this strain was replaced by a neomycin-resistance gene under the control of Lamda Pr promoter (λ Pr-neo) resulting in strain TF8A λ Pr-neo::*Δupp* (Tanaka et al., 2013). The sequence of this region was amplified using λ Pr-neo primer pairs (λ Pr-neo1: CGCTACGCCAGAACACCGATTGAGCTGAT; λ Pr-neo2:CTTAGCTCCTGAAAATC TCGGATCC). The resulted PCR product was used for the transformation of BSB1 strain in order to achieve the master strain *B. subtilis* BSB1N (Figure 24), this strain is a prerequisite to carry out markerless deletion.

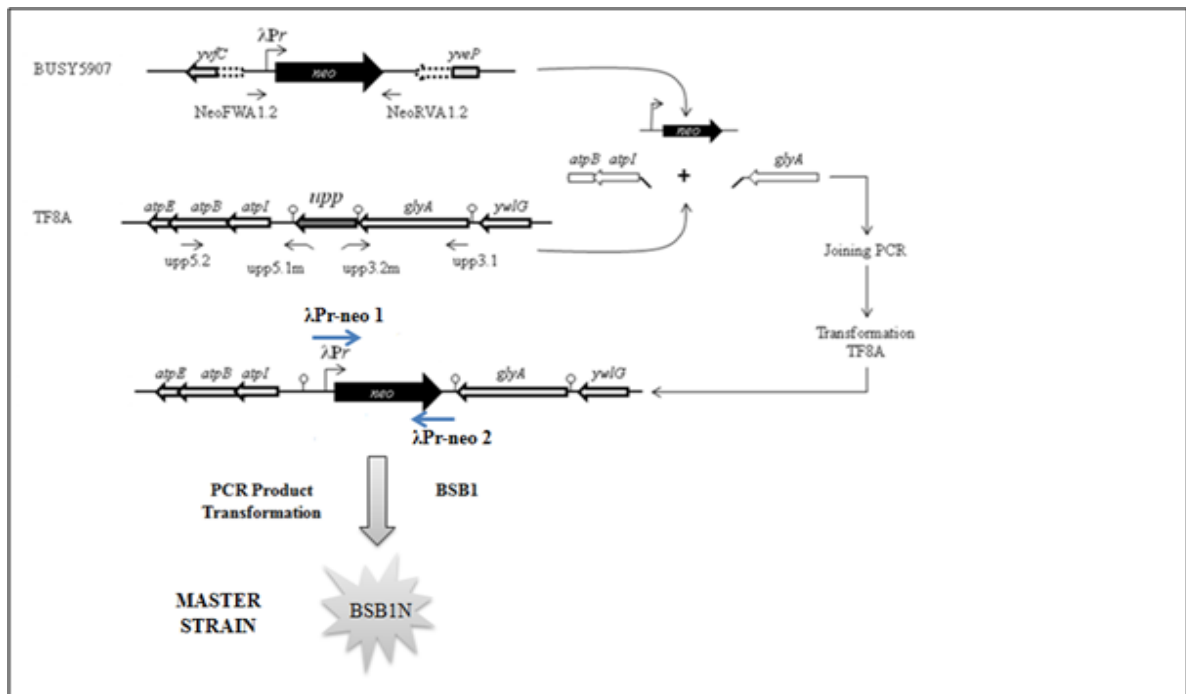


Figure 24 Development of master strain *B. subtilis* BSB1N (Tanaka et al., 2013)

1.3 *Escherichia coli* JM109

This strain was used for the amplification and the conservation of the different plasmids. JM109 is a high-efficiency competent strain ($\geq 1 \times 10^8$ cfu/ μ g DNA) for transformations. The genotype is *recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17* (*rK*⁻, *mK*⁺), *relA1*, *supE44*, Δ (*lac-proAB*), [*F'*, *traD36*, *proAB*, *lacIqZ* Δ M15]. It was supplied by Promega.

2 Plasmids

All the plasmids used in the study are stated below as well as in APPENDIX I.

2.1 Plasmid with *sfp* gene construct

The plasmid containing the gene sequence (*amyE* – *sfp* – *Cm^R* – *amyE* – *Spec^R*) was constructed by the digestion of pBC16 (containing the *sfp* gene) with *EcoRI*. Subsequently, pDG1661 Δ *PvuII* circularised was digested with *EcoRI* for its linearization. The former and later digested fragments were then ligated together to obtain pBG400 (Figure 25). The plasmid concentration was amplified using JM109.

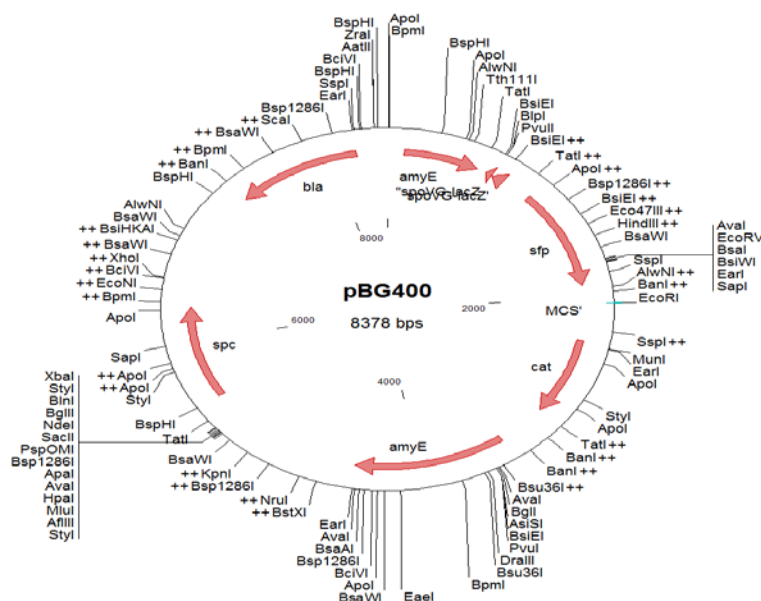


Figure 25 The integration plasmid pBG400 used for the integration of *sfp* gene

2.2 K7010 plasmid

The plasmid K7010 (Figure 26A) contains the cassette used for the construction of markerless deletion strains. This plasmid was derived from pUC19 plasmid. The cassette includes *upp* (uracil phosphoribosyltransferase gene) used previously for counter selection; *phleo* (phleomycin-resistance gene) used as selective marker for cassette integration; *cI* gene (encodes cI protein which acts as a repressor in Lambda phage) was used as counter selection as it can inhibit the expression of λ Pr-*neo* by binding the Lamda promoter (λ Pr).

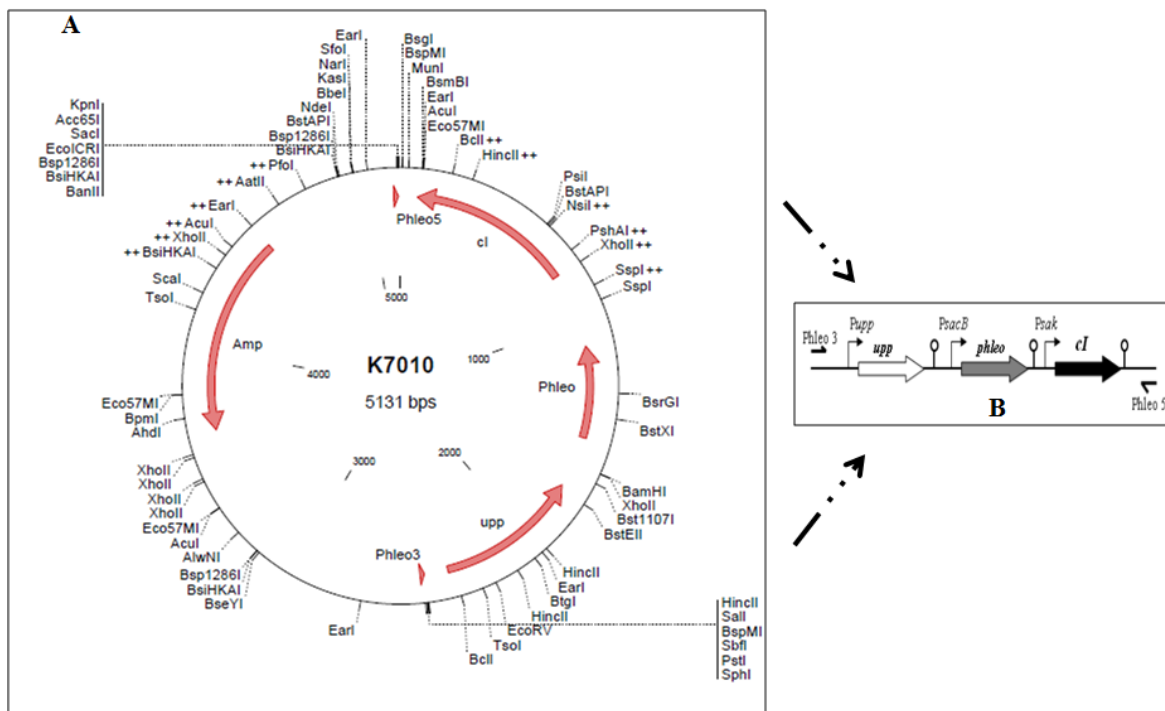


Figure 26 A) Plasmid K7010; B) *upp-phleo-cI* cassette used for markerless deletion

The cassette *upp-phleo-cI* is abbreviated as K7 cassette. The amplification of the cassette (Figure 26B) was achieved using primer pairs (*Phleo5*: 5'-CGACGGCCAGTGAATTCGAGC-3' and *Phleo3*: 5'-AGCTTGCATGCCTGCAGGTCG-3') (Tanaka et al., 2013).

2.3 Plasmid with constitutive promoter P_{repU} -neo construct

The plasmid containing the gene sequence (P_{repU} -neo) was constructed using two plasmids pBG127 and pBG124 (APPENDIX I). The plasmid pBG127 contains *hxlR* sequence while the plasmid pBG124 contains the P_{repU} -neo sequence along with *Hae*III restriction site and *srfAA* sequence. Competent JM109 cells were transformed with pBG124 and pBG127 to obtain EBG132 and EBG128 respectively. Finally, both plasmids (pBG124 and pBG127) were extracted. After the extraction, the plasmids were digested with the following restriction enzymes: pBG127 was digested with *Pae*I and *Bgl*II and pBG124 was digested with *Sph*I and *Bgl*II. The digested samples were purified by agarose gel electrophoresis. The required fragments were extracted from the gel and ligated to obtain as a product the pBG133 plasmid (Figure 27) (Coutte et al., 2010).

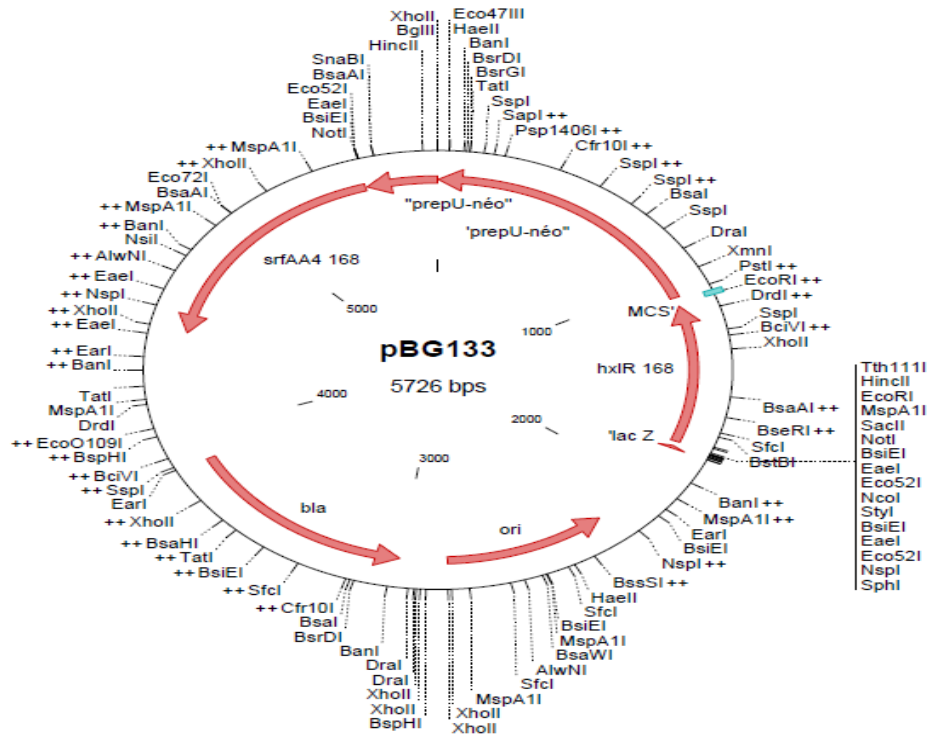


Figure 27 The integration plasmid pBG133 used for the integration of P_{repU} -neo gene

Table 4 Different plasmids and their genotype used in this study

PLASMID	GENOTYPE	SOURCE
pGEM-T Easy	Cloning vector	Promega
pBC 16	Replicative plasmid, <i>sfp</i> gene (B.subtilis ATCC21322), <i>O_γ^{B.subtilis}</i> , Tet ^R	ProBioGem
pDG1661	Integration vector, <i>amyE-spoVG-lacZ-amyE</i> , Ap ^R , Cm ^R , Spc ^R	Guerout-Fleury et al., 1996
pBG400	1.3kb <i>EcoRI</i> fragment of <i>sfp</i> of pBC16 ligated with <i>EcoRI</i> linearized pDG1661, Cm ^R , Spc ^R	This work
K7010	pUC19, <i>upp-phleo-cl</i> , Phleo ^R	INRA, Micalis
pBG124	pGEM-T Easy, <i>HaeIII-srfAA</i> , P _{repU} , Ap ^R , Nm ^R	This work
pBG127	pGEM-T Easy, <i>hxlR</i> , P _{repU} , Ap ^R , Nm ^R	This work
pBG133	1.8kb <i>SpHI-BglII</i> fragment of pBG127 ligated to 3.9kb <i>SpHI-BglII</i> fragment of pBG124, <i>hxlR-PrepU-srfAA</i> , Ap ^R , Nm ^R	This work
pUC57	Cloning vector	GenScript
pBG399	pUC57 ligated to 2.3kb fragment (P _{repU} -partial <i>srfAA</i>), Ap ^R , Erm ^R	This work
pBG401	2.3kb <i>AccI-PmlI</i> fragment of pBG399 ligated to 3.9kb <i>AccI-PmlI</i> fragment of pBG133, Ap ^R , Erm ^R	This work

Table 5 Different *E.coli* strains and their genotype used in this study

STRAIN	CLONED PLASMID	RESISTANCE	SOURCE
<i>E. coli</i> JM109	Competent cells	-	Promega
<i>E. coli</i> EBG108	pDG1661	Ap ^R , Cm ^R , Spc ^R	ProBioGEM
<i>E. coli</i> EBG400	pBG400	Ap ^R , Cm ^R , Spc ^R	This work
<i>E. coli</i> EBG128	pBG127	Ap ^R , Nm ^R	Coutte et al., 2010
<i>E. coli</i> EBG132	pBG124	Ap ^R , Nm ^R	Coutte et al., 2010
<i>E. coli</i> EBG133	pBG133	Ap ^R , Nm ^R	Coutte et al., 2010
<i>E. coli</i> EBG399	pBG399	Ap ^R , Erm ^R	This work
<i>E. coli</i> EBG401	pBG401	Ap ^R , Erm ^R	This work

4 Culture Media

Several culture media have been used in this study. Routine revival of culture and selection of mutants after insertion and deletion have been carried out in conventional LB medium with/without agar (APPENDIX II). Both *B. subtilis* 168 and *B. subtilis* BSB1 are naturally competent strains. In order to carry out modification of strains a competence medium was used to effectively transform the strains (APPENDIX II). Modified version of the classical Landy medium (Landy and Warren, 1948) has been used in which the glutamic acid was replaced with ammonium sulphate. This medium was used for the analysis of surfactin production along with either 0.1M MOPS or phosphate buffer as a buffering agent. Two variants of modified Landy medium were used based on the buffering agent (APPENDIX II). During extracellular metabolite analysis using ¹H-NMR, MOPS can interfere in the signals provided by the metabolites. So MOPS was replaced by K-buffer as a buffering agent. When K-buffer was used it was found that the high concentration of Mg²⁺ interacts with phosphate ions present in K-buffer, so the concentration of magnesium sulphate was reduced in comparison to the other version of modified Landy medium in order to avoid precipitation (APPENDIX II). In addition to modified Landy medium, two different types of media have been also used to study the surfactin production analysis i.e. TSS medium supplemented with 16 amino acids (Brinsmade and Sonenshein, 2011) in order to analyze the impact of glucose concentration as a sole carbon source on surfactin production and AMBER medium in order to analyze the glycolytic mutants as it contains two carbon sources (Glucose and Malate) as well as to have common nutrient environment among all the partners of European project (AMBER) which is focused to determine the linkage between central carbon metabolism and various cellular processes (APPENDIX II). The buffer solution used in former was MOPS and in the latter was K-buffer. Selective media were prepared by adding various antibiotics in LB-agar medium. The following concentration of various antibiotics were added, Chloramphenicol (Cm) 5 µg/ml; Neomycin (Neo) 5 µg/ml, Erythromycin (Erm) 2 µg/ml; Phleomycin (Ph) 4 µg/ml; Spectinomycin (Spec) 100 µg/ml or Tetracycline (Tet) 20 µg/ml for the selection of the desired *B. subtilis* mutant strains where as Erythromycin (Erm) 300 µg/ml; Ampicillin (Ap) 100 µg/ml or Neomycin (Nm) 50 µg/ml for the desired *E. coli* mutant selection. The pH of all culture media was adjusted to 7.0 using 3M KOH/1M H₂SO₄. All the chemicals used for the preparation of the culture media are of analytical grade.

METHODS

5 DNA Manipulation

5.1 DNA extraction

Genomic DNA extraction of *B. subtilis* strains was performed using the Wizard® Genomic DNA Purification Kit (Promega Corp., Madison, WI, USA). The extraction procedure comprises four steps; the first step is the lysis of cells using lysozyme and the RNA is digested by RNase. The cellular proteins are then removed by a salt precipitation step, while leaving the high molecular weight genomic DNA in solution. Finally, the genomic DNA is concentrated and desalted by isopropanol precipitation and dissolved in a TE buffer. The concentration and purity (A_{260}/A_{280}) were measured using NanoDrop (Thermo Scientific NanoDrop Lite Spectrophotometer, USA).

5.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was carried out using the PCR Master Mix (2X) (Thermo Scientific Fermentas, Villebon sur Yvette, France). A PCR mixture was prepared constituted of PCR Master Mix (1X), 0.5 μ M forward and reverse primer, DNA (200 ng) and the volume was made up with MQ water. The reaction was carried out under the following conditions: 5 min at 94°C; (30s at 94°C, 30s at 55°C, X min at 72°C) for 25 cycles; 10 min at 72°C, where X depends on the length of the sequence to be amplified. Generally, the rate of amplification of Taq polymerase is 1kb/minute. After PCR, the desired fragment length was verified using gel electrophoresis on 0.8% agarose gel.

5.3 DNA purification

Recovery of DNA from agarose gels was performed with GeneJET Gel Extraction kit (Thermo Scientific Fermentas). This kit combines the convenience of spin-column technology with the selective binding properties of a uniquely-designed silica-gel membrane. Efficient recovery of DNA was carried out using elution buffer provided by the kit.

5.4 Ligation

Ligation of PCR products with pGEM-T Easy vector (Promega Corp.) was carried out using ligation mixture of pGEM-T Easy vector system. The reaction mix is constituted of ligation buffer (1X), pGEM-T Easy Vector (50 ng), PCR product (X μ l), T4 DNA Ligase (3 Weiss units) and Nuclease-free water to make final volume 10 μ l. The volume of PCR product was determined using following equation:

$$X = \frac{\text{Vector (ng)} \times \text{Insert size (kb)}}{\text{Vector size (kb)}} \times \frac{3}{1}$$

This equation was used for the ligation of inserts into different vectors and for this purpose the DNA Ligation Kit from Takara (Ozyme, Saint Quentin en Yvelines, France) was used. All ligation reactions were incubated overnight at room temperature.

5.5 Plasmid extraction

Plasmid extraction was carried out using GeneJET Plasmid DNAPurification Kit (Thermo Scientific Fermentas). This kit also utilizes spin-column technology and selective binding to silica-gel membrane. The extraction procedure consists in three basic steps: preparation bacterial lysate, adsorption of DNA onto the membrane, washing and elution of plasmid DNA using an elution buffer. The concentration and purity (A_{260}/A_{280}) were measured using NanoDrop (Thermo Scientific NanoDrop Lite Spectrophotometer, USA).

5.6 Restriction digestion

Restriction endonucleases were supplied by Thermo Scientific (Fermentas, Villebon sur Yvette, France). The restriction digestion was carried out at 37°C for 15 min or 1 hr based on FastDigest or conventional restriction endonucleases. This procedure is followed by gel electrophoresis on 0.8% agarose gel along with DNA ladder (O'Gene Ruler Mix, Thermo Scientific) for the analysis of proper product.

In all cases, the instructions of the suppliers were followed to carry out the techniques.

6 Formalization of reaction network

A reaction network was formalized to analyze the metabolic pathway of branched chain amino acids (BCAAs). This reaction network was based on partial kinetic knowledge (Coutte et al., 2015). A reaction network with partial kinetic knowledge consists of a finite set of species, a subset of inflow species, a subset of outflow species, and a finite set of reactions. A reaction is a multiset of pairs between species and roles. Here the following set of identifiers: reactant, product, activator, accelerator, or inhibitors have been used. A species is called a modifier of a reaction if its role is equal to activator, accelerator, or inhibitor. Modifiers are neither consumed nor produced, but the presence of a modifier affects the rate of the reaction. Note that the same species can be used several times in the same reaction and with different roles. The partial knowledge on the kinetics of a reaction is given by the roles that are assigned to its species. The intuition is as follows. The rate of a reaction is increased by reactants, activators, and accelerators, and it is decreased by inhibitors. A reaction is enabled only if all its reactants and activators are present, i.e. have non-zero concentrations, while some of the accelerators may be absent. Steady state semantics was followed in the reaction network.

In a deterministic semantics, a chemical solution is considered as a function of species with non-negative real numbers including zero. For most kinds of species, this number represents the concentration of the species, but in other cases it may also denote the activity of a promoter or the activity of a promoter binding site. Reactions can be applied to chemical solutions. In this case, the reactants are consumed and the products are produced, respectively, as many times as they occur in the reaction. A reaction can only be applied if all its activators are present (have a non-zero concentration), while its accelerators may be absent. The higher the concentration of an accelerator or activator is, the higher will be the rate of the reaction. A reaction is applicable even if some inhibitors are present, but the higher the concentration, the slower will be the rate of the reaction.

Reaction networks are usually situated in a context, which may be adjacent reaction networks, or the “chemical medium” from which some species may inflow and to which some species may outflow during wet lab experimentation. The exchange with the context is modeled by the set of inflow and outflow species of the network: inflow species may inflow from the context, while outflow species may outflow into the context. The precise

rate laws of inflows and outflows are unknown, except the assumption that the outflow rate must increase with the concentration of the outflow species.

Reaction networks in our modelling language are represented as graphs similar to Petri nets (the concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm). These graphs contain two kinds of nodes, round nodes for representing its species and boxed nodes for representing its reactions. More precisely, any species “*S*” is represented by a round node and any reaction with name “*r*” by a boxed node. Solid edges either link a reactant to its reaction, or a reaction to one of its products. There are three kind of dashed edges, which start at the three kinds of modifiers. An accelerator edge links an accelerator to a reaction, an activator edge links an activator to a reaction, and an inhibitor edge links an inhibitor to a reaction. An input edge points from the context to an inflow species *S*, while an outflow edge points from an outflow species *S* to the context. For convenience, a last kind of edges was introduced as a shortcut for a product that is degraded by a hidden reaction. The species nodes were used with three different colors, which indicate their biological roles. The dynamics of the network is not concerned by the colors though. Metabolite is denoted by yellow and in blue proteins. There is a third color for “artificial species” that serve for modelling regulation, such as for instance the promoter of the *ilv-leu* operon. The node colors are irrelevant for the dynamics of the network (Figure 29).

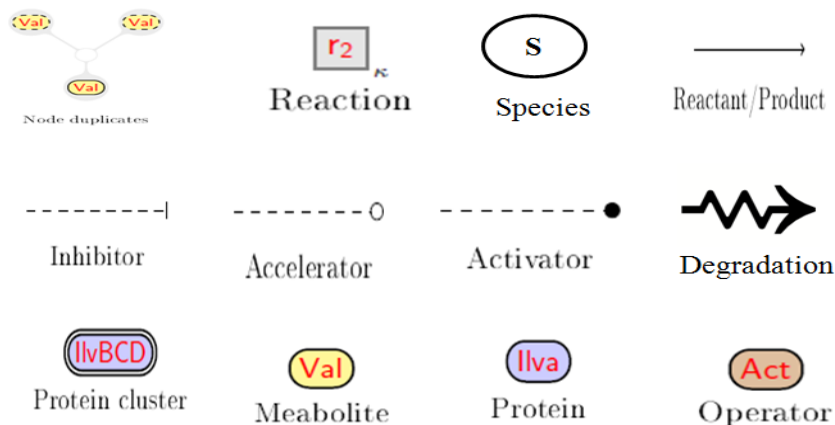


Figure 29 Various identifiers in the reaction network

7 Mutant construction

7.1 Deletion

Based on *B. subtilis* branched chain amino acids reaction network, various mutants were constructed. Mutants were constructed both on the background strain *B. subtilis* 168 and *B. subtilis* BSB1N. The technique (Figure 30) used for the development of markerless mutant strains is known as Pop-In and Pop-Out technique (Tanaka et al., 2013).

Table 6 Primers used for deletion mutant construction

GENES	PRIMER	SEQUENCE
<i>bcd</i>	forward ¹	5'-ACAGCACCCCTTAAGAGCTGGC-3'
	reverse ¹	5'-CGACCTGCAGGCATGCAAGCTATCCAGATTG TTCATCCTGGC-3'
	forward ²	5'-GCTCGAATTCAGTGGCCGTCGCCAGGATGAA CAATCTGGATAACGGCCACAGTGTATTAAGC-3'
	reverse ²	5'-AGCACAATCGCTTCAACTTCGC-3'
<i>codY</i>	forward ¹	5'-GAGGCAATTACGCTTTGGCAG-3'
	reverse ¹	5'-GCTCGAATTCAGTGGCCGTCGAGAAAATC TAAAATCTCATTAA-3'
	forward ²	5'-CGACCTGCAGGCATGCAAGCTTAAATGAGATTTT AGATTTTCGATAAATAATCCTCCTAAACATTC CTCGCTCGAATTCAGTGGCCGTCGAGAAAATCT-3'
	reverse ²	5'-GCTGCAGTTAGAGAGATGCTAG-3'
<i>lpdV</i>	forward ¹	5'-AGAGCTGCAGCGATTGACCG-3'
	reverse ¹	5'-CGACCTGCAGGCATGCAAGCTGACTACGTCATACTCAG TTGC-3'
	forward ²	5'-GCTCGAATTCAGTGGCCGTCGCAACTGAGTAT GACGTAGTCATTTCAACCCGCATCCAACGC-3'
	reverse ²	5'-CTATCTCATCGGACAGCAGGC-3'
<i>tnrA</i>	forward ¹	5'-CAGTACAGCAAATTCAGTGG-3'
	reverse ¹	5'-CGACCTGCAGGCATGCAAGCTCGGACTTTTA TTATTTAACGGTCATTTTCCACCCTGGATG-3'
	forward ²	5'-GCTCGAATTCAGTGGCCGTCGTTAAATAATAAAAGTCCGGC-3'
	reverse ²	5'-AGCATTACACGGTAAAAGACG-3'
<i>yvbW</i>	forward ¹	5'-AAATGATGACTGGAGCTGGC-3'
	reverse ¹	5'-CGACCTGCAGGCATGCAAGCTCCCTTAAATAAACCTG-3'
	forward ²	5'-GCTCGAATTCAGTGGCCGTCGGCGCAGGTTTATTTAAGGGG TCAAAGGACGCAAGCATCAG-3'
	reverse ²	5'-GTCAATTGGCCGCTGTCCAAAG-3'

Two sets of primer pairs were designed for upstream and downstream region of the gene to be deleted (using the sequence from SubtiList World-Wide Web Server and supplied by Eurofins Genomics, Ebersberg, Germany). These two sets of primers were named (forward¹, reverse¹) for the upstream gene and (forward², reverse²) for the downstream gene, with respect to the gene of interest (Table 6). During the design of the primers sets, direct repeat sequence were either added to the primer sequence reverse¹ or forward² as well as complement sequence of 5' and 3' region of the K7 cassette (*upp-phleo-cl*) are also

added to the primer sequence of reverse¹ and forward². PCR was carried out using specific primer pair (forward¹, reverse¹) and (forward²,reverse²) each with 0.5 μM final concentration and 200 ng of master strain chromosomal DNA under standard condition to produce DNA fragments of atleast 1.0 kb long.

The PCR was carried out under the following conditions: 5 min at 94°C; (30s at 94°C, 30s at 55°C, 2 min at 72°C) for 25 cycles; 5 min at 72°C. The PCR products were purified using gel extraction kit (Qiagen, Hilden, Germany). The DNA fragments generated using (forward¹,reverse¹) and (forward²,reverse²) was mixed equally along with *upp-cI-phleo* cassette at a proportion 2:1 and a joining PCR reaction was conducted under the following conditions: 5 min at 94°C; (15 s at 94°C, 15 s at 55°C, 12 min at 65°C) for 12 cycles; (15 s at 94°C, 15 s at 55°C, 12 min at 65°C) for 25 cycles; 10 min at 72°C in the presence of forward¹ and reverse² (0.1 μM final). The resultant product was confirmed through electrophoresis on 0.8% agarose gel and further used to transform competent cells of the master strain (*B. subtilis* BSB1N).

Positive selection of deletion mutants were carried out through phleomycin resistance on LB medium plates, which were incubated at 37°C up to 24 hours. Colonies, with proper cell morphology were streaked twice on the same medium along with on neomycin-LB medium plates in order to purify the colony. Each clone was checked for the absence of the deleted chromosomal region through PCR using primer pair forward¹ and reverse², respectively. Strains constituting the expected deleted chromosomal structure were preserved in glycerol stock at -80°C. These strains with resistance to phleomycin and sensitive to neomycin were known as Pop-In strains.

The Pop-In strain was inoculated in 1 ml liquid LB medium for 6-7 hours at 37°C at 160 rpm. After incubation, the entire culture was lawned on neomycin-LB medium plates, which were incubated at 37°C up to 24 hours. Colonies with proper morphology were streaked on neomycin-LB medium plates, phleomycin- LB medium plates and LB medium plates containing both neomycin and phleomycin and incubated at 37°C up to 24 hours.

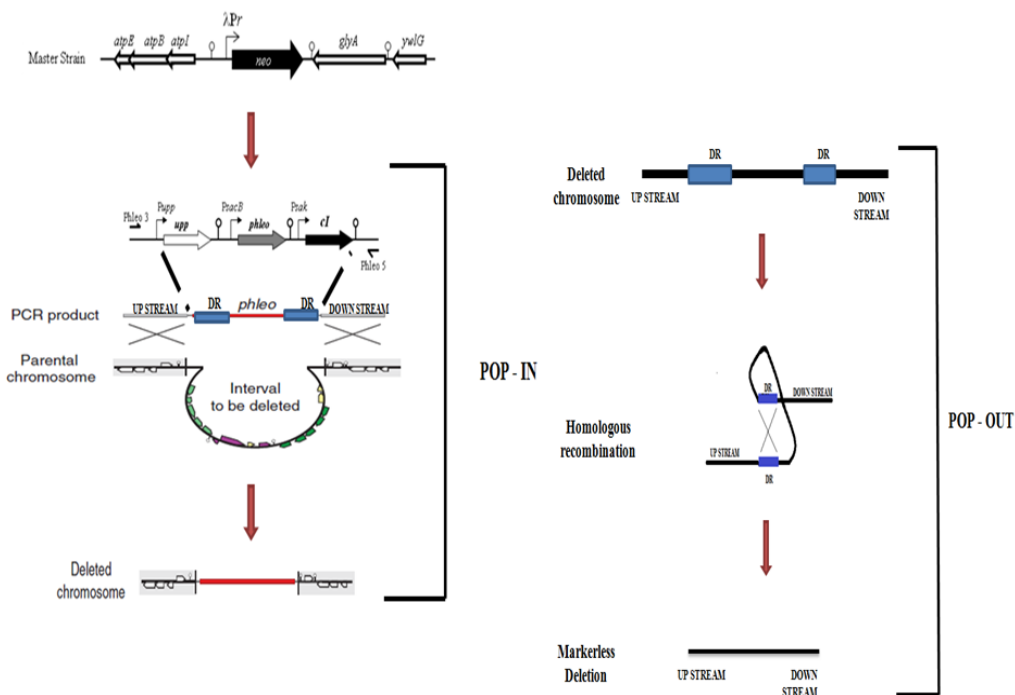


Figure 30 General overview of Pop-In and Pop-Out technique

Colonies which were grown only on neomycin-LB medium plates were treated as positive clones and were later verified through PCR using primer pair forward¹ and reverse² for the excision of the K7 cassette. Strains with no K7 cassette were preserved in glycerol stock at -80°C and were known as Pop-Out strains.

7.2 Insertion

All mutants were transformed with plasmid pBG400 for the integration of *sfp* gene at the *amyE* locus. The transformation protocol has been described in the next section. After transformation, the cells were spreaded on LB-Chloramphenicol plates at 37°C overnight. Each clone obtained were analyzed through replica plate method. LB-Chloramphenicol and LB-Spectinomycin were the chosen selective media. At an event of double cross-over integration, the clones will be resistant only to LB-Chloramphenicol and sensitive to LB-Spectinomycin. The positive clones were also analyzed through PCR using primer pair (*sfp_fw* and *sfp_rev*); the primer sequence is shown in Table 7. Integrated *sfp* gene in the

strains resulted in positive haemolytic activity and negative amylase activity due to disruption of *amyE* locus.

In order to evaluate the role of promoter sequence on surfactin production, the native promoter of *B. subtilis* strain was replaced by a constitutive promoter P_{repU} derived from *S. aureus*. The constitutive promoter *B. subtilis* strains obtained through transformation of the wildtype and derivative strains with pBG401. Upon transformation, they were lawned on LB-Erythromycin plates and incubated at 37°C overnight. Clones obtained were analyzed through PCR using primer pair (P_{repU_fw} and P_{repU_rev}); the primer sequences are shown in Table 7.

Table 7 Primers used for insertion mutant construction

GENES	PRIMER	SEQUENCE
<i>sfp</i>	<i>Sfp_fw</i>	5'-TGACAATCTTTTGCAGACGG-3'
	<i>Sfp_rev</i>	5'-AGGTGTCAAGTTGTTGATGAGC-3'
P_{repU}	P_{repU_fw}	5'-AAGCGGAACCGTAGAAACAAACATGC-3'
	P_{repU_rev}	5'-ATTCCATTGTTGATCCACTCCCTTTTCG-3'

8 Transformation

8.1 *E.coli*

High efficiency *E. coli* JM109 strains were used for the amplification of plasmids. Transformation was performed as described in pGEM[®]-T and pGEM[®]-TEasy vector systems technical manual from QIAGEN.

JM109 cells (50 µL) were transferred to a 15 ml falcon tube in which overnight ligation mixture was added. The tube were flicked gently and placed on ice for 20 min, and then heat-shock was applied to the cells by immersing the tube in water bath at 42°C for 90 secs. The tubes were returned immediately to ice for 2 minutes and after that 900 µL LB medium was added. The entire mixture was incubated at 37°C with shaking for 1.5 hours. After incubation, the entire mixture was spread on selective medium based on plasmid, with which it was transformed.

8.2 *B. subtilis*

Both *B. subtilis* 168 and BSB1 strains are naturally transformable strains. Strain(s) are streaked out on LB-plates (with/without antibiotic based on the strain) overnight at 37°C. A colony was inoculated in 2 ml of completed MC (APPENDIX II). The mixture was incubated at 37°C with shaking for 4-5 hours (or more until the culture is turbid). Then, 400 µL of cells were mixed with 1 µg of DNA (gDNA/plasmid) and it was incubated again at 37°C with shaking for 2 hours. After incubation, the entire mixture was spread on LB-antibiotic plates based on the recipient resistance along with control cells without transformation in different plates.

9 Surfactin analysis

All mutants obtained were analyzed for surfactin production in order to determine the impact of gene for surfactin overproduction in comparison to control. The *sfp* gene required for lipopeptide production was inserted in all mutants prior to surfactin analysis. This gene can promote surfactin production by converting the Apo-synthetase to holo-synthetase.

9.1 Culture conditions

Firstly, single colony of mutant strain(s) was inoculated in LB-antibiotic medium. It was incubated at 37°C with shaking conditions. Next, required volume of culture was centrifuged at 8,000 rpm for 10 min. The cell pellets were washed twice to remove residual medium. Then, it was inoculated in the experimental media at 37°C with 160 rpm overnight to acclimatize the nutrient environment. Finally, the overnight culture was centrifuged at 8,000 rpm for 10 min and the cells were inoculated in the experimental medium. In all experiments, the inoculation OD₆₀₀ was maintained 0.2 and carried out at 37°C with variable shaking conditions. The surfactin analysis was carried out in both Biolector[®] device and shake flask condition.

Biolector[®] (Mp2-labs GmbH; Baesweiler; Germany) is a high-throughput fermentation system (Figure 31A) available on REALCAT platform of the University of Lille-Sciences and Technologies. Cultures were carried out in 48 wells flower plate (Figure 31B)

designed for the Biolector[®]. Each well consists of probes for the measurement of pH, dissolved oxygen, fluorescence and biomass (Figure 31C). The dry weight of each strain was determined through calibration curve in Biolector[®] with a slope of 0.0307, offset value of -0.5323 and a coefficient of regression of 0.9944.



Figure 31 A) Biolector[®] fermentation system; B) 48-well MTP; C) Probes inside each well

Experiments were carried out with 1 ml culture volume with a shaking frequency of 1100 rpm, under high oxygenation rate of $55 \text{ mmol l}^{-1}\text{h}^{-1}$ with a shaking diameter of 3 mm along with humidity control. A real-time monitoring of growth of the mutants was carried out using BioLecton V.2.4.1.0 software. Experiments were carried out in triplicate and two biological replicates were performed. Results are expressed as mean value with standard deviation

Conventional Erlenmeyer culture flask of 500 ml and 1000 ml with 10% filling ratio were used to carry out the shake flask experiment, as described previously (Coutte et al., 2010).

In both conditions, at every time point, sample was transferred into 1.5 ml sterile eppendorf tubes. The samples were centrifuged at 8,000 rpm for 10 mins, the supernatant

were transferred into new sterile 1.5 ml eppendorf tubes for surfactin analysis and the sampling of the same time-point was carried out in triplicate.

9.2 Surfactin analysis by RP-HPLC

Surfactin was analyzed using reverse phase C18 HPLC (600 s, Waters, USA) equipped with a C18 column (250 mm x 3.0 mm, Grace, USA). It was quantified using surfactin standard with 98% purity (Lipofabrik, Villeneuve d'Ascq, France). Surfactin was eluted during 20 min under an isocratic condition (80:20) of ACN/TFA (0.1% TFA in 100% ACN) and H₂O/TFA (0.1% TFA in 100% ACN) at a flow rate of 0.6 ml/min. The chromatogram at 214 nm was extracted and the peaks were selected on the basis of certain curve of secondary derivative at 214 nm with Savitzky-Golay smoothing of 15. The surfactin concentration was normalized each time with the dry weight of the strain at that particular time interval. Results are expressed as mean values with standard deviation and the statistical analysis of the data has been made using one-way ANNOVA with Tukey's multiple comparison test (p-value <0.0001).

10 LC-MS-MS analysis of surfactin isomer

Samples were analyzed by reverse phase UPLC-MS (UPLC, Waters, Acquity class H) coupled with a single quadrupole MS (SQDetector, Waters, Acquity) on an Acquity UPLC BEH C18 (Waters) 2.1 × 50 mm, 1.7 μm column. A method based on acetonitrile gradients that allowed the simultaneous detection of all three LP families. Elution was started at 30% acetonitrile (flow rate of 0.60 ml min⁻¹). After 2.43 min, the percentage of acetonitrile was brought up to 95% and held until 5.2 min. Then, the column was stabilized at an acetonitrile percentage of 30% for 1.7 min. Compounds were identified on the basis of their retention times compared with authentic standards (98% purity, Lipofabrik, Villeneuve d'Ascq, France) and the masses detected in the SQDetector. Ionization and source conditions were set as follows: source temperature, 130°C; desolvation temperature, 400°C; nitrogen flow, 1000 l h⁻¹; cone voltage, 120 V. This technique was carried out in Gembloux Agro-Bio Tech, University of Liege, Belgium.

11 Extracellular metabolome

At each time point, 2 ml of sample were filtered through 0.45 µm pore size filter (Sarstedt AG, Nürnberg, Germany) to obtain cell-free medium for the analysis of extracellular metabolome. All samples were stored at -20°C prior to analysis.

During the analysis, 400 µl of sample was transferred to 5 mm glass tube (length 7 inch; NORELL ST-500, NORELL, USA) and the sample was buffered to pH 7.0 by adding 200 µl of 0.2 mol/l sodium hydrogen phosphate buffer solution which was prepared with 50% D₂O (Euriso-Top, St-Aubin Cedex, France), this solution provides a NMR-lock signal. In addition, the buffer solution also contains 1 mol/l TSP (3-trimethylsilyl-[2, 2, 3, 3-D₄]-1-propionic acid) (Sigma-Aldrich, USA) as an internal standard.

The analysis was carried out using ¹H-NMR and the samples were measured at 600.27 MHz and a temperature of 310K. The spectrum was obtained utilizing Bruker AVANCE-II 600 NMR spectrometer and was controlled by TOPSPIN 2.1 software (Bruker Biospin GmbH, Germany). Identification and quantification of NMR spectra was carried out using AMIX-Viewer v3.9.11 software (Bruker Biospin GmbH, Germany). Both standard compounds (Sigma-Aldrich, USA) and Human Metabolome Database (HMDB) was applied for the identification of spectra while the quantification was carried out by exploiting the signal area of Quant Ref signal at 15 ppm (Kohlstedt et al., 2014b). This technique was carried out in Ernst-Moritz-Arndt-University Greifswald, Institute of Biochemistry, Germany.

CHAPTER 3: RESULTS AND DISCUSSION

1. Is intracellular pool of leucine a limiting factor for surfactin biosynthesis?

B. subtilis BBG111 is a mutant of *B. subtilis* 168 able to produce two types of lipopeptides i.e. surfactin and fengycin. Coutte *et al.* previously showed that the deletion of *pps* gene which is responsible for the biosynthesis of fengycin leads to an increase of surfactin biosynthesis (Coutte *et al.*, 2010). One of the hypotheses suggested by the authors to explain this result was a better availability of precursors required for surfactin biosynthesis as both the lipopeptides share several precursors. In addition, the replacement of the native promoter (P_{srfA}) of surfactin operon with a constitutive promoter (P_{repU}) did not lead to a high enhancement of surfactin production. Hence, it was again speculated that there might be limitation of precursor molecules, which are required for biosynthesis of surfactin.

Surfactin molecule contains seven amino acids within its peptide moiety. Out of the seven, five positions are occupied by branched chain amino acids i.e. valine and leucine. It was thus hypothesized that an increase in the intracellular pool of leucine pool will have a positive impact on the surfactin production as the molecule contains four molecules of leucine (Figure 33).

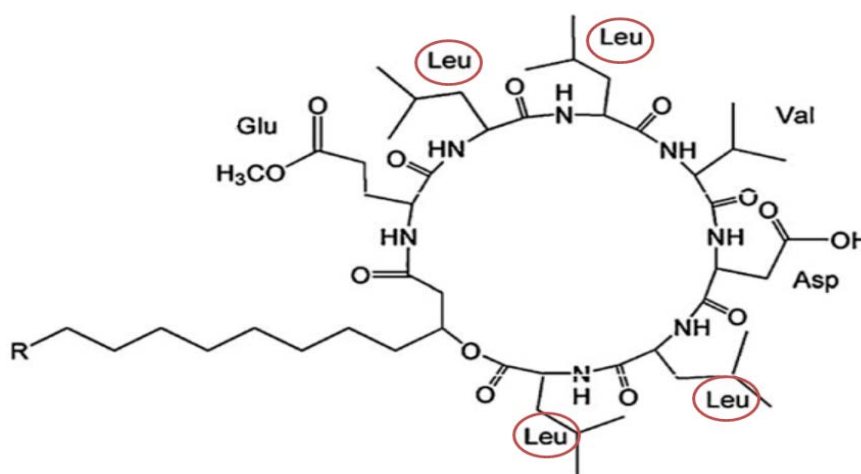


Figure 32 General surfactin structure: It consists of four leucine molecules among seven amino seven amino acids present within the peptidic moiety, the position of the leucine molecules are encircled

In order to investigate the hypothesis that an increase in the intracellular pool of leucine can lead to an increase in the surfactin production, the strain *B. subtilis* BBG111 (Coutte et al., 2010) was cultivated in the presence of L-leucine. Two different cultures were carried out in Erlenmeyer flask at 37°C and an agitation of 160 rpm (materials and method: Surfactin analysis). The first culture medium used was Landy (Landy and Warren, 1948) with ammonium sulphate (instead of glutamic acid) and the second chosen nutrient environment was Landy medium with ammonium sulphate and L-leucine (instead of glutamic acid). The Landy medium with ammonium sulphate instead of glutamic acid will be called modified Landy medium momentarily.

Samples were withdrawn after 6 and 24 hours of incubation. Biomass and surfactin analysis were carried out (materials and method: Surfactin analysis). Cultures were performed in triplicate and results are expressed as mean value with standard deviation in Figure 34.

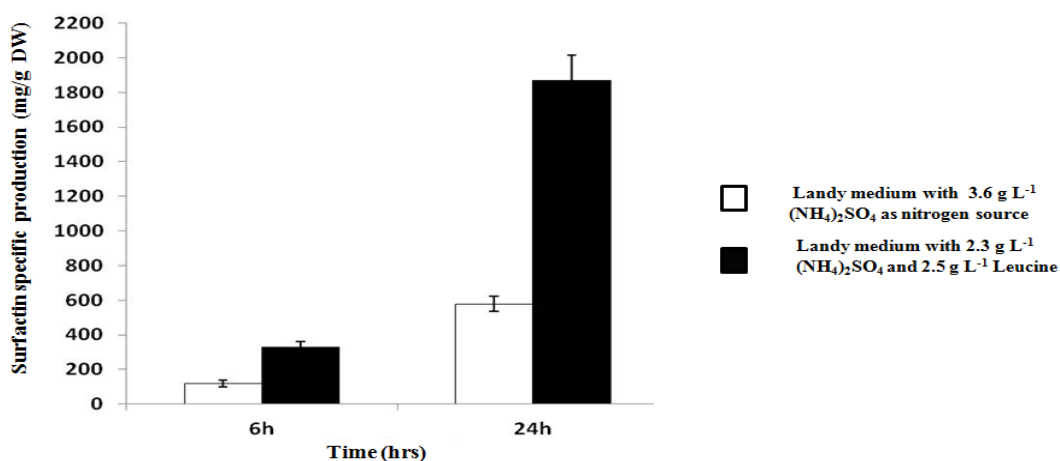


Figure 33 Leucine feeding : Impact on specific surfactin yield with *B. subtilis* BBG111 in two different media condition with and without the presence of leucine in the culture media at 37°C in shake flask condition with 160 rpm. Results are expressed as mean value with standard deviation.

When the culture was grown in modified Landy medium with only ammonium sulphate as the nitrogen source, specific surfactin production was 100 mg/g of biomass dryweight (DW) and 600 mg/g of DW at 6 h and 24 h of culture. However, upon addition of L-leucine (2.5 g/l) and (NH₄)₂SO₄ at a concentration of 2.3 g/l, the specific surfactin

production was enhanced 3-folds at 6 h and 24 h, respectively in comparison to the former condition. The maximum specific surfactin production of 1800 mg/g of DW was obtained in Landy medium supplemented with 2.3 g/l $(\text{NH}_4)_2\text{SO}_4$ and 2.5 g/l of L-leucine for 24 h cultivation (Figure 33). Thus, it can be inferred from the results that the hypothesis raised at the beginning of the study was found to be true as it highlights the impact of leucine feeding on surfactin biosynthesis.

Hence, leucine as a monomer of surfactin forms an essential component for the enhancement of its production. The leucine intracellular pool results from leucine transporter efficiency on one side and from the anabolic pathway on the other side.

The metabolic flux analysis is a central method of cell engineering to increase the productivity of host microorganisms such as bacteria, yeast and fungi. In this area, many fundamental studies have been conducted on the pathway regulations and metabolic flux analysis (Otero and Nielsen, 2010). The use of bioinformatics in recent years helped to develop tools for reconstruction of metabolic pathways and genome scale metabolic modelling in order to overexpress metabolites, but most of them are only applied to primary metabolites and not on secondary metabolites (Almaas et al., 2004) (Ranganathan et al., 2010) (Goelzer et al., 2008) (Sohn et al., 2010). In this work, the main emphasis is an overproduction of secondary metabolites, i.e. the surfactin biosynthesized by NRPS. To overcome the surfactin precursor limitation, here for the first time an original and fully integrated approach of synthetic biology to overproduce surfactin was developed (in collaboration with the Biocomputing team of CRISAL laboratory of Lille).

2. How to enhance leucine production within the intracellular pool ?

2.1 Formalization of the model of BCAA metabolic pathway

Branched chain amino acids (BCAA) include isoleucine, valine and leucine. These amino acids are regulated in a similar way. A metabolic network for the production of these amino acids is presented in Figure 34.

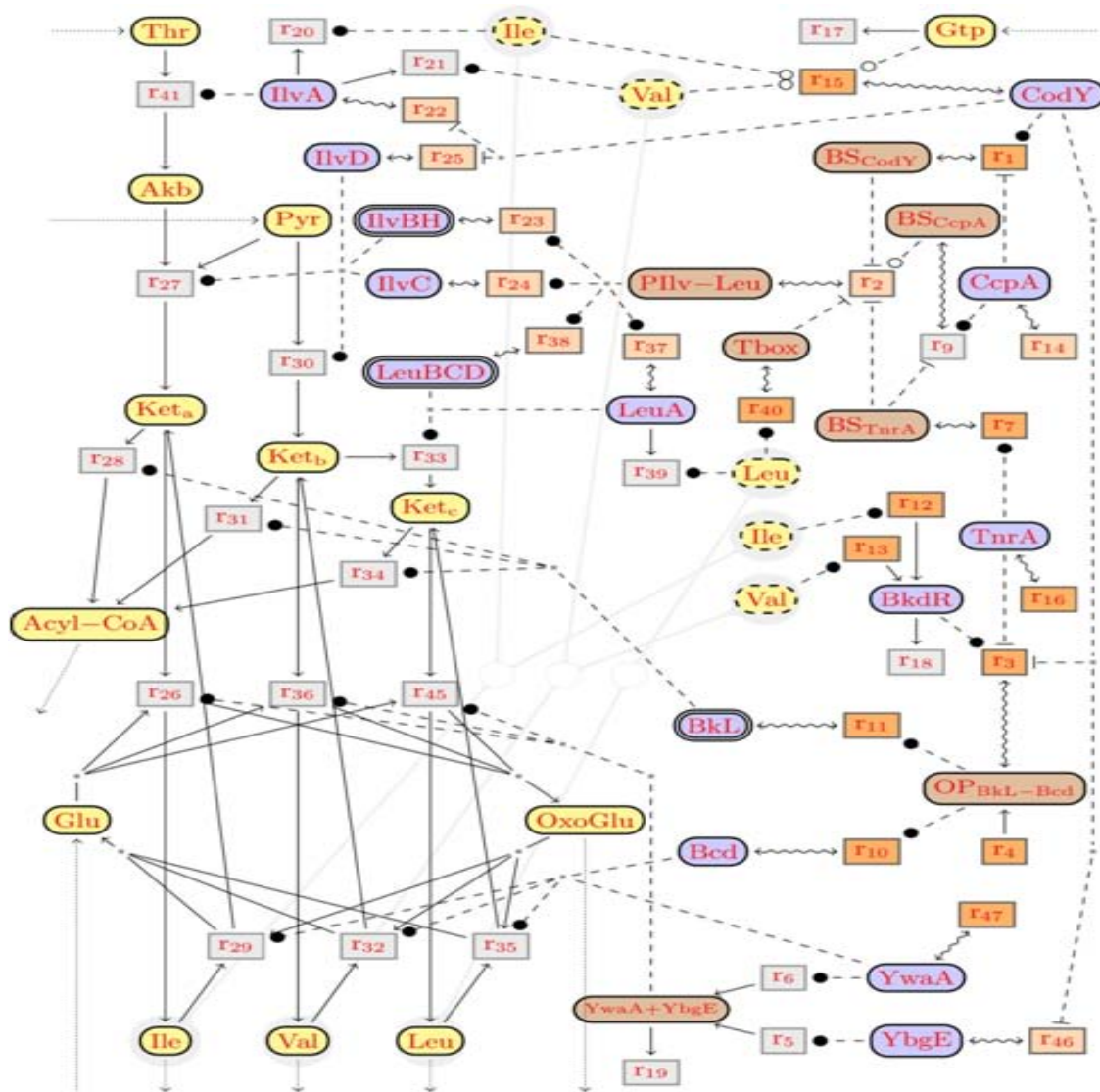





Figure 34 Branched chain amino acid reaction network. The concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm.

The Figure 34 describes the metabolic transformations of Thr, Pyr and Akb into Ile, Val, and Leu, concomitantly with its regulation. This model mainly follows the abstraction level chosen by the previous work in (Goelzer et al., 2008) (Mäder et al., 2012), but the regulation is expressed more cleanly within this formal modeling language.

As the precise kinetics is unknown, the complex nature of the regulation of the inhibitors and the activators in the reaction were extracted based on comprehensive literature study and assumption. Although few detailed information regarding all inhibitors and activators were not available but the reaction network was formalized using this partial knowledge. The information regarding the metabolites, proteins and actors of this model and all the reactions have been provided APPENDIX III.

In order to group several proteins that participate in the same reaction, artificial species were introduced that were called protein clusters, and were represented as for instance . This protein cluster presents all the proteins generated by the *bkL* operon and in a similar fashion the proteins LeuB, LeuC and LeuD as well as IlvB and IlvH are represented in clusters LeuBCD and IlvBH, respectively.

Split-points were used to link different copies of the same species. For instance, one copy

of Leu, which was drawn as  was linked together with the original  over a split point. Logically, there is no difference between a species and its copies; it just allows nicer graphs with fewer intersections of edges. Edge-clusters were used to represent several groups with the same sources or targets (for instance, the activator edges from BkL and YwaA to r_{35} , r_{32} , and r_{29} are clustered).

Production of Isoleucine, Valine, and Leucine For clarity, the intermediates that produce the outflows of the pathway, i.e. Ket_a for Ile in r_{26} , Ket_b for Val in r_{36} , and Ket_c for Leu in r_{45} were considered. All three reactions must be activated by YwaA/YbgE, that is either by protein YwaA (r_6) or protein YbgE (r_5), which are exchangeable catalysts following (Thomaides et al., 2007). There are also inverse reactions r_{29} , r_{32} , r_{35} , but with different activators, the proteins Bcd and YwaA. The intermediates Ket_a , Ket_b , and Ket_c can be transformed alternatively into Acyl-CoA by reactions r_{28} , r_{31} , and r_{34} , which can then flow out into the context for fatty acid biosynthesis. This is regulated by protein BkL.

The creation of Ile, Val, and Leu from Ket_a , Ket_b and Ket_c is always coupled with a

parallel transformation of Glu into OxoGlu, and the inverse decompositions are coupled with the inverse transformation of OxoGlu into Glu. The required Glu for the production of Ile, Val, and Leu, can be input from the glutamate biosynthesis cycle of the context (derived from OxoGlu) produced OxoGlu can flow out from this pathway to the context (as for example used in the TCA cycle). The intermediates are produced as follows: Ket_a is produced from Akb by r_{27} , Ket_b from Pyr by r_{30} , and Ket_c from Ket_b by r_{33} . Akb can be produced from Thr by reaction r_{41} , which in turn can also be input from the threonine biosynthesis pathway of the context (which is linked to the glycolysis pathway via the aspartate biosynthesis). Pyr can be input from the glycolysis pathway of the context. IlvA activates the conversion of Thr into Akb. IlvBH (a cluster of IlvB and IlvH), IlvC and IlvD activate the transformation of Akb to Ket_a. The proteins LeuA and LeuBCD (a cluster of LeuA, LeuB, and LeuC) perform a sequence of reactions justifying the transformation of Ket_b to Ket_c (Mäder et al., 2012). The genes of the proteins IlvBH, IlvC, LeuA, and LeuBCD are co-located in the same operon (*ilv-leu*), so they are under the dependence of the same single promoter P_{Ilv-Leu} (Tojo et al., 2005). This specific regulation is described below.

Regulation The expression of the different enzymes involved in the production of branched chain amino acids (explained above) are denoted by the following reactions: r_{22} for IlvA; r_{25} for IlvD; r_{23} for IlvBH; r_{24} for IlvC; r_{38} for LeuBCD; r_{37} for LeuA; r_{11} for BkL; r_{10} for Bcd; r_{47} for YwaA and r_{46} for YbgE. All these expression reactions are controlled by a complex regulation (Brinsmade et al., 2010)(Molle et al., 2003)(Mäder et al., 2004b). The proteins CcpA, CodY, TnrA, and BkdR have regulatory functions and are expressed by the reactions r_{14} , r_{15} , r_{16} and r_{12} , r_{13} , respectively. The expression reactions of these regulators can be modulated by different factors from outside the context, which are not modeled here. GTP comes from the context and can be degraded through the reaction r_{17} . This metabolite increases the affinity of CodY for its binding sites, as well as Ile and Val (Shivers and Sonenshein, 2004) (Tojo et al., 2005) (Handke et al., 2008) (Molle et al., 2003) (Mäder et al., 2004b).

Expression of IlvA, IlvD, and YbgE are dependent on their own promoters and their expressions are down-regulated by CodY at the transcriptional level (Molle et al., 2003) (Mäder et al., 2004b), which is represented by reactions r_{22} , r_{25} , and r_{46} , respectively. Hatfield and Umbarger (1970) it was further showed that IlvA is deactivated by both Ile and Val (reactions r_{20} , r_{21}). The transcription of all proteins IlvC, IlvBH, LeuA, and

LeuBCD starts through the activation of their promoter $P_{\text{IIV-Leu}}$, which controls the reactions r_{23} , r_{24} , r_{37} and r_{38} , respectively (Mäder et al., 2004b) (Brinsmade et al., 2010). Its regulation is regulated by the reaction r_2 and is dependent on several mechanisms, which are described before. In addition to what has been described above, the down-regulation of the promoter $P_{\text{IIV-Leu}}$ by Leu is also contributed by a ribosome-mediated attenuation mechanism Tbox (r_{40}) (Brinsmade et al., 2010) (Grandoni et al., 1992) and the deactivation of LeuA by Leu, which is represented by the reaction r_{39} . Proteins BkL and Bcd lie under the action of two common but independent promoters (Prm_b and Prm_c), whose activity level is represented by actor $\text{OP}_{\text{BkL-Bcd}}$. Among them one is constitutive (no regulation, represented by the reaction r_4), and the other is positively impacted by BkdR and down-regulated by both TnrA and CodY, as reflected by reaction r_3 (Michel Debarbouille et al., 1999). Although not regulated on the transcriptional level, protein BkdR is activated by Ile and Val, which is modeled by introducing the two BkdR producing reactions r_{12} and r_{13} . Proteins YbgE (expressed by the reaction r_{46}) is down regulated by CodY. However, no regulation is known for protein YwaA, which is expressed by the reaction r_{47} .

2.2 Knockout prediction

The knockout prediction algorithm was applied to the reaction network (Figure 34). The integrated approach was focused on a modelling language for reaction networks with partial information and on a new knockout prediction algorithm based on abstract interpretation that applies to all models written in this language (Niehren et al., 2016). There were two types of prediction provided, firstly, regarding the flux of various species or metabolites (s) and secondly, regarding single knock-out of genes. Both predictions were made based on increasing the production of leucine ($y_{\text{Leu}} = \uparrow$). In this context, $x_s = \downarrow$ denotes a decrease in the flux of species or metabolite (but not to zero) and $y_s = \uparrow$ denotes an increase in the outflux of species or metabolite (but not from zero).

In all the flux solutions (Table 8) it was observed that $x_{\text{Glu}} = y_{\text{oxoGlu}} = y_{\text{Leu}} = \uparrow$, so it means that any increase of y_{Leu} requires increasing the influx of Glu and outflux of OxoGlu.

The prediction based on flux does not require any reaction knockout within this context (Figure 34). In addition to these context changes, few others were also predicted which include $y_{\text{Acyl-CoA}} = \downarrow$, i.e. the outflux of Acyl-CoA is decreased, or else that $x_{\text{Pyr}} = \uparrow$, so that the influx of Pyr is increased. Both possibilities are plausible, since for overproducing Leu, either the production of Leu by r_{45} must be increased, and this requires more Pyr, or else the outflux of Leu into the fatty acid pathway (via reactions r_{35} and r_{34}) must be decreased, and this imposes a higher outflux of Acyl-CoA.

All solutions corresponding to single knockouts have been provided in Table 9. This can be done automatically by the tools developed by the Biocomputing team. In addition, an informal explanation of the positive effect of these single knockouts with respect to leucine overproduction has also been provided (Table 9).

Table 8 Solutions related to flux where no knockouts are required

SOLUTION	x_{Glu}	x_{Pyr}	$y_{\text{Acyl-CoA}}$	y_{Leu}	y_{oxoGlu}
1	↑	↑	~	↑	↑
2	↑	~	↓	↑	↑

Table 9 Single knockout predictions for leucine overproduction

KNOCKOUT	DESCRIPTION OF REACTION	KNOCKOUT EFFECT	MUTANT
r ₁	CodY binding to BS _{CodY}	increases P _{Ilv-Leu} activity	BBG251
r ₃	bind BkdR to BkL-Bcd promoter	decrease Bcd and thus r ₃₅	
r ₄	constitutive expression of BkL-Bcd operon	decreases Bcd and thus r ₃₅	
r ₇	TnrA binding to BS _{TnrA}	increases P _{Ilv-Leu} activity	
r ₁₀	Bcd expression	deactivates reaction r ₃₅	BBG253 and BBG259
r ₁₁	BkL expression	deactivates Ket _a , Ket _b , Ket _c outflow via Acyl-CoA by r ₂₈ , r ₃₁ , r ₃₄	BBG255 and BBG261
r ₁₂	activate BkdR by Ile	decreases speed of r ₃ and thus Bcd	
r ₁₃	activate BkdR by Val	decreases speed of r ₃ and thus Bcd	
r ₁₅	CodY expression	increase P _{Ilv-Leu} activity	BBG254 and BBG260
r ₁₆	TnrA expression	increase P _{Ilv-Leu} activity	BBG256 and BBG262
r ₄₀	Tbox of Leu attenuation	increase P _{Ilv-Leu} activity	BBG252
r ₄₇	expression of YwaA	increase speed of r ₃₅	
r ₁₄ r ₄₆	expression of CcpA expressions of YbgE	unclear unclear	

These effects can be seen in the model rather easily. It turns out, however that knockout of reactions r₁₄ and r₄₆ do not have a clear positive effect. The only effect of a knockout of reaction r₄₆ is the deactivation of r₄₅ producing Leu, which is unwanted. The knockout of reaction r₁₄ has a negative effect on the activity of P_{Ilv-Leu}. This is not very plausible for leucine overproduction: it could still be argued that a decrease of the activity of P_{Ilv-Leu} will decrease the speed of r₂₇ which in turn will increase the speed of r₃₀, but at the same time, it also has a negative effect on the activation of r₃₀, which counter balances this positive effect on the Leu production. Moreover, it has been demonstrated previously that a knockout of *ccpA* causes a decrease of *ilv-leu* expression (Brinsmade et al., 2010).

This new knockout prediction algorithm was applied for the overproduction of intracellular leucine production, which in turn might increase the production of surfactin in different mutants of *B. subtilis* 168 or *B. subtilis* BSB1 strain.

2.3 Mutant constructed

Previous section provided various single knockout candidates (Table 9) by abstract interpretation for the overproduction of leucine based on the reaction network (Figure 34). Based on the prediction, selected few genes which regulated the expression of *ilv-leu* operon were deleted in order to validate the prediction results and evaluate its effect on surfactin production.

The mutants were constructed both with *B. subtilis* 168 and *B. subtilis* BSB1, in order to check the efficacy of the prediction results based on *B. subtilis* strains.

The deletion was carried out in the model organism *B. subtilis* BSB1 strain using PopIn – PopOut technique (Figure 30). This technique produced entire removal of the gene, which was crucial to study their impact on intracellular leucine pool and eventually on surfactin production. After deletion, the gene of interest was knockout entirely, which can be verified through PCR verification using forward¹ and reverse² primer pairs (Table 6). The band size obtained was 1982 bps, 3075 bps, 1953 bps and 1819 bps for the deletion of *bcd* (lane 1), *codY* (lane 2), *lpdV* (lane 3) and *tnrA* (lane 4) genes, respectively (Figure 35).

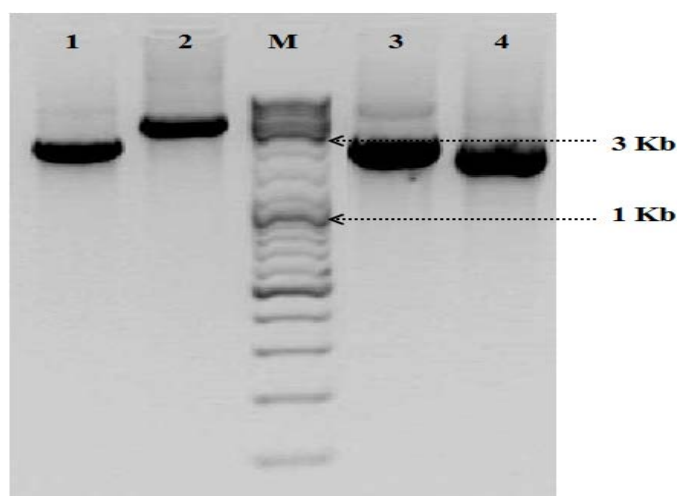


Figure 35 PCR verification for markerless deletion **1:** Δbcd ; **2:** $\Delta codY$; **3:** $\Delta lpdV$; **4:** $\Delta tnrA$; **M:** 1 kb Ladder

Genomic DNA of PopIn mutant strains were extracted and was used for the transformation of BBG111 (*B. subtilis* 168 derivative strain) through natural transformation method.

Table 10 Deletion mutants derived from *B. subtilis* 168 strain based on prediction results of the branched chain amino acid reaction network

STRAIN	GENOTYPE	SOURCE
BBG111	BS168, <i>sfp</i> ⁺ : cm ^r	Coutte et al., 2010
BBG251	BS168, <i>sfp</i> ⁺ : cm ^r , Δ <i>codY</i> site II	This work
BBG252	BS168, <i>sfp</i> ⁺ : cm ^r , Δ <i>t-box</i>	This work
BBG253	BS168, <i>sfp</i> ⁺ : cm ^r , Δ <i>bcd</i> : ph ^r	This work
BBG254	BS168, <i>sfp</i> ⁺ : cm ^r , Δ <i>codY</i> : ph ^r	This work
BBG255	BS168, <i>sfp</i> ⁺ : cm ^r , Δ <i>lpdV</i> : ph ^r	This work
BBG256	BS168, <i>sfp</i> ⁺ : cm ^r , Δ <i>tnrA</i> : ph ^r	This work

The mutants with desired gene deletion were selected on LB-agar containing phleomycin (4 μ g/ml). Two other strains SRB94 and SRB87 were obtained from Sonenshein Lab (Brinsmade et al., 2010). SRB94 was generated through site directed mutagenesis in the region II of CodY binding site at *ilv-leu* promoter, while SRB87 is a mutant strain devoid of stem-loop terminator structure responsible for leucine-dependent transcriptional regulation; *sfp* gene was inserted in the former and latter strain to generate BBG 251 and BBG 252. The mutants with *B. subtilis* 168 as background strains are provided in Table 10.

B. subtilis BSB1 strain is also devoid of *sfp* gene, which is essential for the conversion of apo-synthetase to holo-synthetase and results in the production of lipopeptides. This strain was transformed with pBG400 using natural transformation method for the insertion of *sfp* gene in the *amyE* locus and positive mutants were selected as described in material and methods. The *sfp* gene was inserted in all the markerless mutant strains obtained through PopIn – PopOut technique. Double knockout strains were also obtained through the transformation of genomic DNA of PopIn mutant strain into PopOut mutant strain and simultaneously selected on selective LB-agar plates. The mutant strains were verified through PCR. The mutants with *B. subtilis* BSB1 as background strain are provided in Table 11.

Table 11 Deletion mutants derived from *B. subtilis* BSB1 strain based on prediction results of the branched chain amino acid reaction network

STRAIN	GENOTYPE	SOURCE
BBG258	BSB1, <i>sfp</i> ⁺ : cm ^r , neo ^r	This work
BBG259	BSB1, <i>sfp</i> ⁺ : cm ^r , Δbcd , neo ^r	This work
BBG260	BSB1, <i>sfp</i> ⁺ : cm ^r , $\Delta codY$, neo ^r	This work
BBG261	BSB1, <i>sfp</i> ⁺ : cm ^r , $\Delta lpdV$, neo ^r	This work
BBG262	BSB1, <i>sfp</i> ⁺ : cm ^r , $\Delta tnrA$, neo ^r	This work
BBG265	BSB1, <i>sfp</i> ⁺ : cm ^r , $\Delta tnrA$, $\Delta codY$: ph ^r	This work

Thus, the markerless mutants developed along with BBG 111 and BBG 258 were treated as control strain for *B. subtilis* 168 and *B. subtilis* BSB1 strain for growth and surfactin analysis, respectively. The knockout strategy used provided complete deletion of the gene of interest. The markerless technique adopted here was found to be an efficient tool to carry out multiple deletion in one strain without antibiotic limitation.

According to the prediction results (Table 9) based on the model, various genes were deleted. Firstly, the CodY binding to its high affinity site on the P_{Ilv-Leu} promoter (r₁) and secondly, the *codY* expression (r₁₅), the strains BBG251 and BBG254 were constructed in *B. subtilis* 168 strain while only the latter was constructed in *B. subtilis* BSB1 strain (BBG260). TnrA also negatively regulates the *ilv-leu* expression under nitrogen-limited conditions, the knockout predictions giving as a result the deletion of the TnrA binding on the P_{Ilv-Leu} promoter (r₇) and the expression of *tnrA* (r₁₆). To achieve the same effect, the expression of *tnrA* (r₁₆) was completely suppressed in the mutant strain BBG256 and BBG262. Both the expression of *codY* (r₁₅) and *tnrA* (r₁₆) were suppressed simultaneously in one mutant strain BBG265. Leucine intracellular concentration can inhibit the expression of *ilv-leu* operon via Tbox attenuator. In the mutant strain BBG252, Tbox was suppressed (r₄₀), as recommended by the predictions. Lastly, prediction on leucine degradation was avoided by the knockout of the gene *bcd* (r₁₀), giving the strains BBG253 and BBG259 as well as deletion of *lpdV* (r₁₁), a gene within the *bkL* operon producing strains BBG255 and BBG261. The strains developed through this strategy were used for

surfactin analysis. It was observed in Figure 34, that both TnrA and CodY binds to the upstream region of *ilv-leu* operon and repress its activity and this deletion was carried out to evaluate the role of double mutants in surfactin through this network so that the results can be used for the refinement of the model.

2.4 Growth and surfactin analysis

The mutants obtained with both *B. subtilis* 168 and *B. subtilis* BSB1 along with control strains were cultivated in high-throughput system of fermentation Biolector[®] (Mp2-labs GmbH; Baesweiler; Germany) available on REALCAT platform of University Lille 1. Cultures were performed in Landy medium with ammonium sulphate at 37°C with shaking frequency of 1100 rpm. During the analysis of *B. subtilis* 168 and its derivative strain, the medium was supplemented with tryptophan at 16 mg/L. Samples were withdrawn when the cells were in exponential phase (about 6 h of culture) as described in (material and method-Surfactin analysis). Indeed, at this time the *ilvB* operon (Nicolas et al., 2012) is highly expressed and surfactin measurement was used as an indirect indicator to evaluate the impact of these genetic modification on the intracellular pool of leucine.

2.4.1 Growth analysis

Growth kinetics study of all the mutant strains were carried out during the exponential growth phase. No significant differences were observed for the growth rate of the two control strains (BBG111 and BBG258). The specific growth rate was found to be $0.47 \pm 0.06 \text{ h}^{-1}$ and $0.42 \pm 0.01 \text{ h}^{-1}$ for BBG111 and BBG258, respectively. The specific growth rates of various mutants were different from the control strains BBG111 and BBG258 as shown in Table 12.

Table 12 Specific growth rate of all mutant strains derived from BBG111 and BBG258 in Landy medium with ammonium sulphate

STRAINS	SPECIFIC GROWTH RATE (h ⁻¹)
BBG111 (BS168, <i>sfp</i> ⁺ : cm ^f)	0.47 ± 0.06
BBG251 (BS168, <i>sfp</i> ⁺ : cm ^f , $\Delta codY$ site II)	0.51 ± 0.01
BBG252 (BS168, <i>sfp</i> ⁺ : cm ^f , $\Delta t\text{-}box$)	0.51 ± 0.02
BBG253 (BS168, <i>sfp</i> ⁺ : cm ^f , Δbcd : ph ^f)	0.39 ± 0.05
BBG254 (BS168, <i>sfp</i> ⁺ : cm ^f , $\Delta codY$: ph ^f)	0.48 ± 0.03
BBG255 (BS168, <i>sfp</i> ⁺ : cm ^f , $\Delta lpdV$: ph ^f)	0.54 ± 0.04
BBG256 (BS168, <i>sfp</i> ⁺ : cm ^f , $\Delta trnA$: ph ^f)	0.47 ± 0.12
BBG258 (BSB1, <i>sfp</i> ⁺ : cm ^f , neo ^f)	0.42 ± 0.01
BBG259 (BSB1, <i>sfp</i> ⁺ : cm ^f , Δbcd , neo ^f)	0.5 ± 0.02
BBG260 (BSB1, <i>sfp</i> ⁺ : cm ^f , $\Delta codY$, neo ^f)	0.46 ± 0.01
BBG261 (BSB1, <i>sfp</i> ⁺ : cm ^f , $\Delta lpdV$, neo ^f)	0.39 ± 0.03
BBG262 (BSB1, <i>sfp</i> ⁺ : cm ^f , $\Delta trnA$, neo ^f)	0.5 ± 0.02
BBG265 (BSB1, <i>sfp</i> ⁺ : cm ^f , $\Delta trnA$, $\Delta codY$: ph ^f)	0.48 ± 0.03

Most of the regulatory genes involved in the branched chain amino acid metabolic pathway are pleiotropic regulators in *B. subtilis*, so the effect on the growth of the strains in which these genes were deleted has also been studied. Growth analysis of the different mutant strains shows that the deletions of *trnA* and *Tbox* have no significant negative impact on the specific growth rate of the strains. These data suggest that no significant influence was contributed by these regulators on the growth of the mutants. The disruption of high affinity CodY binding site and the *codY* deletion has no negative effect on the growth of the strains. Regarding the gene involved in the branched chain amino acid degradation, *bcd* and *lpdV* which code for branched chain amino acid dehydrogenase enzymes (Debarbouille et al., 1999) showed no growth defect too. Deletion may lead to a disturbance in the synthesis of branched-chain fatty acids, which plays an important role in maintaining the fluidity of the membrane lipids (Kaneda, 1991) but this was not observed

during growth analysis. All the obtained mutants were viable because these genes are not listed as essential for the cell (Commichau et al., 2013).

2.4.2 Surfactin analysis

BBG111 and its derivative mutant strains shown in Table 10 (page no. 89) were evaluated for surfactin production.

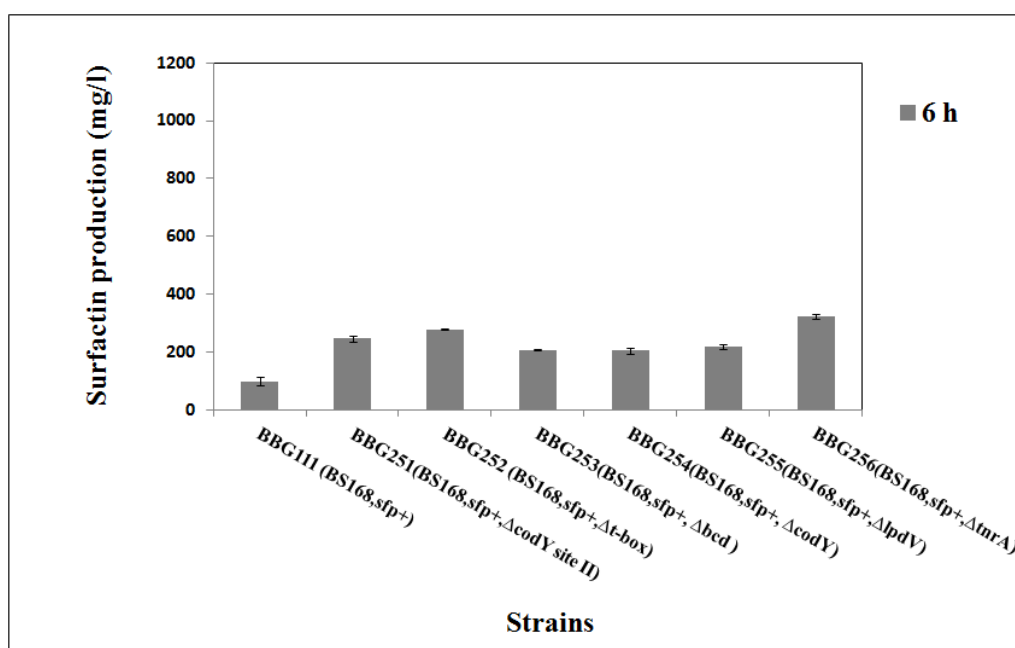


Figure 36 Surfactin production (mg/l) of BBG111 and its derivative strains (*B. subtilis* 168) in Landy medium with ammonium sulphate at 37°C in Biolector with shaking frequency of 1100 rpm. Results are expressed as mean value with standard deviation.

It can be observed from Figure 36, that all the mutant strains produced surfactin with varied concentrations. At 6 hours, the production of surfactin (mg/l) by the mutant strains was higher in comparison to control strain (BBG111). The maximum surfactin production of 324±9 mg/l was obtained with BBG256 (*BS168, sfp⁺, ΔtnrA*). Although the figure does not represent the normalization concentration with dry weight of specific strain, this representation has been provided in the later section.

BBG258 was developed through the insertion of *sfp* gene in *B. subtilis* BSB1 strain. This strain (treated as control strain) along with its derivative mutant strains shown in Table 11 were evaluated for surfactin production.

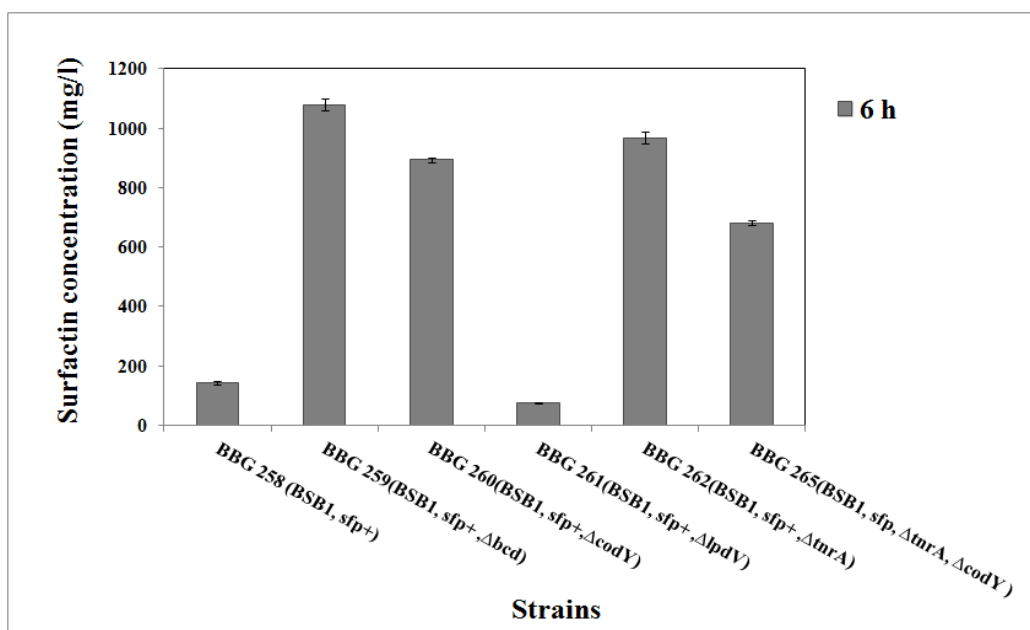


Figure 37 Surfactin production (mg/l) of BBG258 and its derivative strains (*B. subtilis* BSB1) in Landy medium with ammonium sulphate at 37°C in Biolector with shaking frequency of 1100 rpm. Results are expressed as mean value with standard deviation.

All the mutant strains produced different concentration of surfactin, which can be seen in Figure 37. In comparison to control (BBG258) all the mutant strains produced higher surfactin at 6 hours except BBG261. BBG259 (BSB1, *sfp*⁺, Δ*bcd*) produced maximum surfactin with a concentration of 1079±21 mg/l. The normalized representation of surfactin production has been provided in the later section.

Thus, it can be deduced from Figure 36 and Figure 37 that *sfp* was properly integrated into the strains, leading to surfactin producing strain and the deletion has a role in the enhancement of surfactin production through an increase in leucine intracellular pool.

The surfactin concentration (mg/l) produced was normalized by the dryweight (g/l) of each strain at 6 hours to determine the specific surfactin yield by each mutant strain. The specific surfactin yield of BBG111 was found to be 194±32 mg of surfactin per g of dry weight (DW) of cell biomass, whereas for BBG251, BBG252, BBG253, BBG254, BBG255 and BBG256 was found to be 310±16, 366±5, 634±36, 1300±243, 580±58, 818±59 mg of surfactin per g of DW of cell biomass (Figure 38). The maximum surfactin

yield was obtained BBG254 (BS168, *sfp*⁺, $\Delta codY$), which is 7-fold higher in comparison to the control (BBG111).

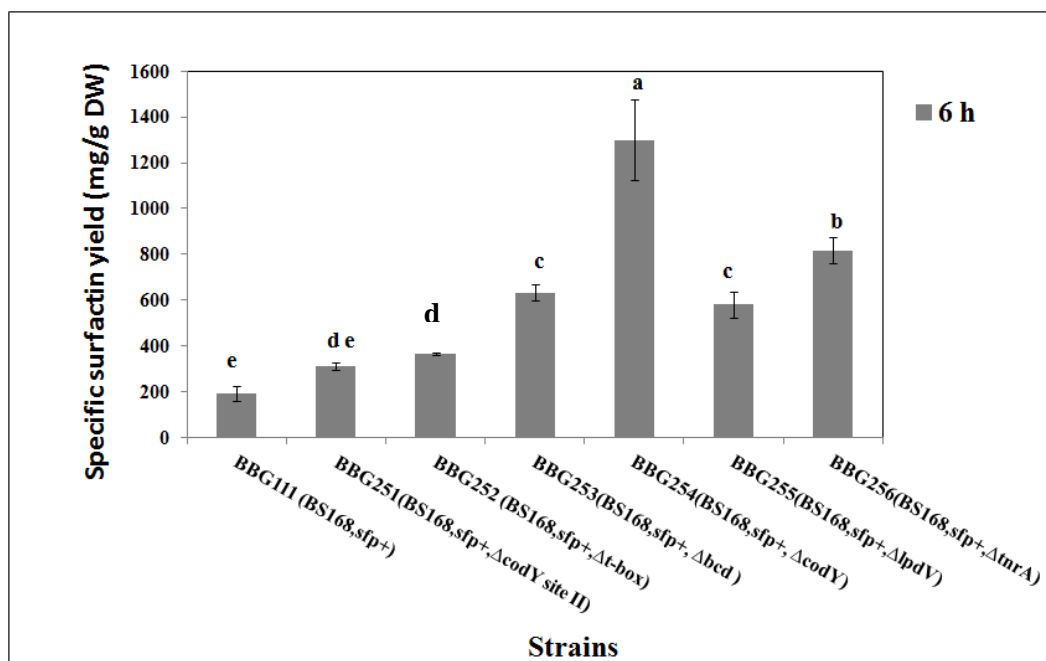


Figure 38 Specific surfactin yield (mg/g DW) of BBG111 and its derivative strains in Landy medium with ammonium sulphate. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey’s multiple comparisons test.

In Figure 36, BBG256 (BS168, *sfp*⁺, $\Delta tnrA$) produced 323.9±9.44 mg/l, that was maximum in comparison to other mutant strains but upon normalization the surfactin yield is found to be 4 times higher than control. It can be observed (Figure 38) that the strains BBG254, BBG256, BBG255 and BBG253 have significant impact on surfactin yield in comparison to control.

When similar experiment was performed with BBG258 and its derivative strains, the specific surfactin yield of BBG258 was found to be 252±24 mg of surfactin per g of cell DW. The surfactin yield for BBG259, BBG260, BBG261, BBG262 and BBG265 was found to be 1158±48, 1277±10, 148±4, 1076±23 and 973±11 mg of surfactin per g of cell dry weight, respectively (Figure 39). The specific surfactin yield of BBG260 (BSB1, *sfp*⁺, $\Delta codY$) was found to be 5 times more than the control strain (BBG258) and a minimum

surfactin yield of 3.8 times in comparison to control was found with BBG265 (BSB1, *sfp*⁺, $\Delta codY$, $\Delta trnA$). In Figure 37, the strain BBG259 produced the maximum surfactin concentration but upon normalization the specific yield was found to be 4.6 times higher than the control.

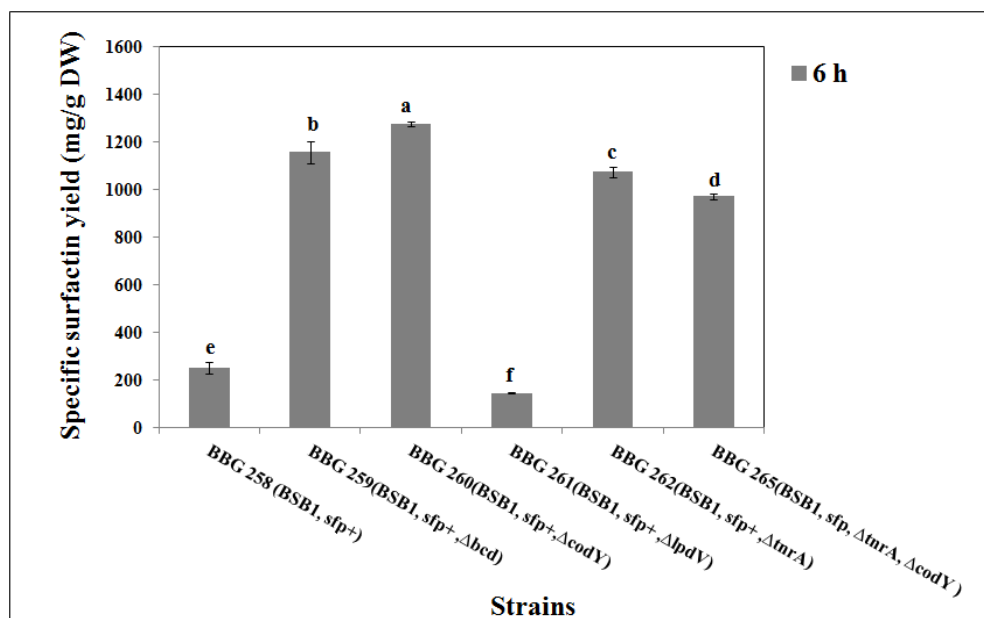


Figure 39 Specific surfactin yield (mg/g DW) of BBG258 and its derivative strains in Landy medium with ammonium sulphate. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test.

It can be observed (Figure 39) that the strains BBG260, BBG259, BBG262 and BBG265 have significant impact on enhancement of surfactin yield in comparison to control.

Thus, the mutant strains constructed with both *B. subtilis* 168 and *B. subtilis* BSB1 strains according to prediction results yielded in an increase in specific surfactin production from 1.6 to 7-folds more in comparison to the respective control strains (BBG111 and BBG258). It was suggested that the strains *B. subtilis* 168 and *B. subtilis* BSB1 are similar and the only difference is that the former is tryptophan auxotroph and the latter is tryptophan prototroph. Genetic studies revealed that there is no difference in the genome level but the analysis of specific surfactin yield of strains with *bcd* and *lpdV* deletion indicates that there might be some difference in the proteome level.

The expression of *ilv-leu* operon is regulated by CodY in the presence of branched chain amino acid and is known to respond to high content of GTP present in exponentially growing phase (Ratnayake-Lecamwasam et al., 2001). In the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, there was an increase of seven times and five times in the specific surfactin production with BBG254 and BBG260 (*codY* deleted; r_{15}), which reached an exceptional value of 1.3 g and 1.27 g of surfactin produced per gram of dried biomass. This provides an evidence for higher expression of *ilv-leu* in the mutant strains, thus, results in an increase of intracellular pool of leucine.

The production of the strain BBG251 (BS_{CodY} deleted; r_1) was not significantly different in this condition compared to the control strain. In this case, it can be supposed that CodY remains active and was able to repress the *ilv-leu* operon via its other binding sites. Indeed, the $P_{\text{Ilv-Leu}}$ promoter contains four binding sites for CodY, in the promoter-proximal CodY-I+II (nt -84/-32) site that is a high-affinity binding site of CodY, while the promoter-distal CodY-III (nt-154/-107) and CodY-IV (185/-168) sites are low-affinity ones (Shivers and Sonenshein, 2004). This shows that modification of CodY regulation in the cell led to an enhancement of specific surfactin production by increasing the production of leucine via the overexpression of *ilv-leu* operon. This observation has been perfectly predicted (Table 9) in the model (Figure 34) that CodY has an impact on the promoter $P_{\text{Ilv-Leu}}$ through the reaction r_{15} . The $P_{\text{Ilv-Leu}}$ is also subject to a negative regulation by TnrA through its binding to the specific sites. This stimulates DNA binding inhibiting transcription initiation and repressing the *ilv-leu* expression (Fujita et al., 2014). To avoid this regulation, the prediction suggested deletion of input of TnrA (r_{16} or to suppress its binding on the $P_{\text{Ilv-Leu}}$: r_7). This repression was disrupted by the deletion of *tnrA* in the strain BBG256 and BBG262. In Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, where the concentration of organic nitrogen was lower, a 4 times increase in specific surfactin production was observed with both strains. These results are quite logical since the impact of this regulator is bestowed only in organic nitrogen limited conditions (Fisher, 1999).

When both the regulatory genes *tnrA* and *codY* were deleted in the strain BBG265, an enhancement of only 3.8 times in specific surfactin production was observed. The production was lower than the one obtained after single knockout of *codY* and almost equal to single knockout of *tnrA*, respectively. Glutamate is the best source of nitrogen utilized by *B. subtilis*. Moreover in the absence of glutamate, ammonium is utilized as a nitrogen source. This is then used for the synthesis of glutamine and glutamate by

glutamine synthetase and glutamate synthase, respectively. *B. subtilis* requires NrgA for the transport of ammonium and NrgB is essential for the expression of *nrgAB* operon during ammonium assimilation (Detsch and Stülke, 2003). TnrA can upregulate the expression of *nrgAB* (Wray and Fisher, 2011). So it can be hypothesized that in BBG254 and BBG260 the production was due to high *ilv-leu* expression, which is supported by proper transport of ammonium ions for nitrogen metabolism required for surfactin synthesis. In BBG256 and BBG262, the production is lower due to lack of ammonium transport but there is still production due to relieve of repression of TnrA on $P_{Ilv-Leu}$. But the specific production with BBG265 ($\Delta codY$, $\Delta tnrA$) was lower than *codY* single mutants due to lack of nitrogen availability and almost equal to *tnrA* single mutants due to removal of repression on *ilv-leu* promoter. The result of the double knockout mutant suggests that the prediction algorithm is limited to the branched chain amino acid residue metabolism, which can predict for single knockout but do not ensure the enhancement for double knockout mutants. This result can also be explained by the fact that the version of prediction algorithm used here can predict multiple knockout without antagonistic effect but cannot accurately predict knockout having synergistic effect.

Intracellular concentration of leucine also downregulates the *ilv-leu* operon using a feedback loop through an attenuator (Grandoni et al., 1992). This 80-bp leucine-dependent stem loop structure (Tbox) has been identified by the predictions as a target to be deleted (r_{40}). The effect of this deletion on specific surfactin production by the BBG252 strain was very low in the Landy medium supplemented with $(NH_4)_2SO_4$ (1.5-fold increase) in comparison to BBG111. Due to absence of leucine in the extracellular medium the regulation does not occur in the Landy medium containing only $(NH_4)_2SO_4$ as nitrogen source. Brinsmade et al. found that the copies of *ilv-leu* transcript were 19 times higher in Tbox mutant (SRB87) in comparison to the wild type strain, when leucine was present in the medium (Brinsmade et al., 2010).

The predictions obtained from the network (Figure 34) were related to the genes involved in the leucine degradation: *ywaA* via r_{47} , *bkl* via r_{11} and *bcd* via r_{10} , and the mechanism that activates the expression of the two latter genes, i.e. expression of BkdR via Ile and Val activation via r_{12} and r_{13} and its binding on the promoter via r_3 or the constitutive expression of the promoter by r_4 . All these knockout predictions have the same effect of interrupting the two stages of reactions that allow the degradation of branched chain amino acids to make fatty acid precursors. To validate the prediction, *bcd* and *lpdV* were deleted

seperately in the strains BBG253 and BBG259; BBG255 and BBG261, respectively. Here, *ywaA* was not chosen because it is also involved in leucine production through the reaction r_{45} . However, YwaA action in this step might be able to be offset by YbgE as these two enzymes are exchangeable and have the same function (Thomaides et al., 2007). In Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, the effect of knockout on *bcd* led to an enhancement of three times and four times in specific surfactin production, respectively. But the deletion had no impact in case of BBG261 whereas an enhancment of three times was observed with BBG255.

A complete model of the most complex and regulated metabolic pathway of *B. subtilis*, i.e. the branched chain amino acid metabolic pathway has been developed. When applied to the model of BCAA metabolic pathway (Figure 34), the prediction algorithm returned 12 plausible single knockout predictions. From these 12 predictions, five mutants were constructed in *B. subtilis* 168 and three similar mutants were constructed in *B. subtilis* BSB1 strain to study the consistency of the prediction. Since, the peptide moiety of surfactin contains four leucine residues; the surfactin concentration was measured as an indicator to detect the intracellular level of leucine. All these results discussed above confirm the efficiency and the relevancy of the modelling language and the abstract interpretation algorithm developed for gene knockout prediction. Indeed in the tested medium conditions, all the predicted knockouts led to an increase in surfactin production. In summary, this work opens numerous perspectives, firstly, to improve the models and the algorithm and secondly, to apply this original approach to more complex secondary metabolites biosynthesis or more complex metabolic pathways such as may be found in eukaryotic organisms.

2.5 Publication and Valorization of these results

ARTICLE

Coutte, F., Niehren, J., Dhali, D., John, M., Versari, C., & Jacques, P. (2015). **Modeling leucine's metabolic pathway and knockout prediction improving the production of surfactin, a biosurfactant from *Bacillus subtilis***. *Biotechnology journal*, 10(8), 1216-1234.

ORAL COMMUNICATION

Jacques, P., Coutte, F., Bechet, M., Guez, J.S., Leclere, V., Dhali, D., Fickers, P., John M., Niehren, J. **Directed Biosynthesis of Antifungal and Biosurfactant Lipopeptides from *Bacillus subtilis***, 10th European Symposium on Biochemical Engineering Science, Lille, France, 8-10 September 2014.

Coutte, F., John, M., Dhali, D., Béchet, M., Niehren J., Jacques, P. **Synthetic biology for the overproduction of lipopeptides in *B.subtilis***, 3rd International Symposium on Antimicrobial Peptides, Lorient, France, 4-6 June 2014.

Coutte.F, Jacques.P, Dhali.D, Guez. J.S, Leclère.V, Auger.S, **Engineering metabolic pathways of *Bacillus subtilis* to modifying pattern of bioactive lipopeptides**, 11th European Symposium on Biochemical Engineering Science, Dublin, Ireland, 11-14 September 2016.

Dhali.D , Coutte.F , Niehren.J, John.M, Versari.C, Jacques.P **Modeling BCAA's metabolic pathway and knockout prediction improving the production of surfactin, a biosurfactant from *Bacillus subtilis***, BioMicroWorld 2015, Barcelona, 28-30 October 2015.

POSTER COMMUNICATION

Dhali.D , Coutte.F , Niehren.J , Versari.C, John.M, Bidenko.V, Auger.S, Jacques.P, **Role of transcriptional regulators on surfactin overproduction : A bottom-up approach**, 8th International Conference on Gram-Positive Microorganisms, Montecatini Terme, Italy, 21-25 June 2015.

3. FACTORS INFLUENCING SURFACTIN PRODUCTION

Surfactin production depends on various factors which include nutrient composition, aeration and genotype of the strain used for production as discussed before in the literature review section (Role of various factors- page no. 45). In this chapter, different experiments have been performed to study the impact of surfactin production in *B. subtilis* 168 and *B. subtilis* BSB1 strains.

3.1 Role of nutrient composition

In order to study the impact of nutrient composition, the strains (Table 10 page no. 89) were cultivated in high-throughput system of fermentation Biolector[®]. Cultures were analysed in TSS medium supplemented with 16 amino acids with shaking frequency of 1100 rpm. Both media were supplemented with tryptophan at 16 mg/L. Samples were withdrawn when the cells were in exponential phase (about 6 h of culture).

Growth kinetics study of all the mutant strains were carried out during the exponential growth phase. The specific growth rate of BBG111 was found to be $0.47 \pm 0.06 \text{ h}^{-1}$ and $0.55 \pm 0.05 \text{ h}^{-1}$ in Landy medium and TSS medium, respectively. The specific growth rates for BBG251, BBG252, BBG253, BBG254, BBG255 and BBG256 in Landy medium were found to be $0.58 \pm 0.06 \text{ h}^{-1}$, $0.55 \pm 0.02 \text{ h}^{-1}$, $0.38 \pm 0.08 \text{ h}^{-1}$, $0.39 \pm 0.03 \text{ h}^{-1}$, 0.54 ± 0.04 and $0.59 \pm 0.01 \text{ h}^{-1}$, respectively. Whereas in TSS medium, the specific growth rates for BBG251, BBG252, BBG253, BBG254, BBG255 and BBG256 were found to be $0.51 \pm 0.01 \text{ h}^{-1}$, $0.51 \pm 0.02 \text{ h}^{-1}$, $0.39 \pm 0.05 \text{ h}^{-1}$, $0.48 \pm 0.03 \text{ h}^{-1}$, 0.58 ± 0.07 and $0.47 \pm 0.12 \text{ h}^{-1}$, respectively (Table 13). The specific growth rate of BBG254 increased in Landy medium in comparison to the specific growth rate in TSS medium whereas for the other strains the medium composition did not affect the growth rate.

Table 13 Specific growth rate of BBG111 and its derivative strains in Landy medium with ammonium sulphate and TSS medium supplemented with 16 amino acids

STRAINS	SPECIFIC GROWTH RATE (h ⁻¹)	
	LANDY MEDIUM WITH AMMONIUM SULPHATE	TSS + 16 AMINO ACID MEDIUM
BBG111 (<i>BS168,sfp⁺:cm^f</i>)	0.47 ± 0.06	0.55±0.05
BBG251 (<i>BS168,sfp⁺:cm^f,ΔcodY site II</i>)	0.51 ± 0.01	0.58 ± 0.06
BBG252 (<i>BS168,sfp⁺:cm^f, Δt-box</i>)	0.51 ± 0.02	0.55 ± 0.02
BBG253 (<i>BS168,sfp⁺:cm^f, Δbcd:ph^f</i>)	0.39 ± 0.05	0.38 ± 0.08
BBG254 (<i>BS168,sfp⁺:cm^f, ΔcodY:ph^f</i>)	0.48 ± 0.03	0.39 ± 0.03
BBG255 (<i>BS168,sfp⁺:cm^f,ΔlpdV:ph^f</i>)	0.54±0.04	0.58±0.07
BBG256 (<i>BS168,sfp⁺:cm^f,ΔmrA:ph^f</i>)	0.47 ± 0.12	0.59 ± 0.01

The decline in specific growth rate of BBG254 in TSS medium might be due to decrease in guanine nucleotide synthesis and CodY is guanine nucleotide sensor (Ratnayake-Lecamwasam et al., 2001). So, in the absence of CodY in minimal media the level guanine nucleotide affects the growth rate of this mutant strain.

All the strains were evaluated for surfactin production in TSS medium. It can be observed from Figure 40, that all the strains produced surfactin with varied concentrations. At 6 hours, the production of surfactin (mg/l) by the mutant strains was higher except for BBG256 in comparison to control strain (BBG111).

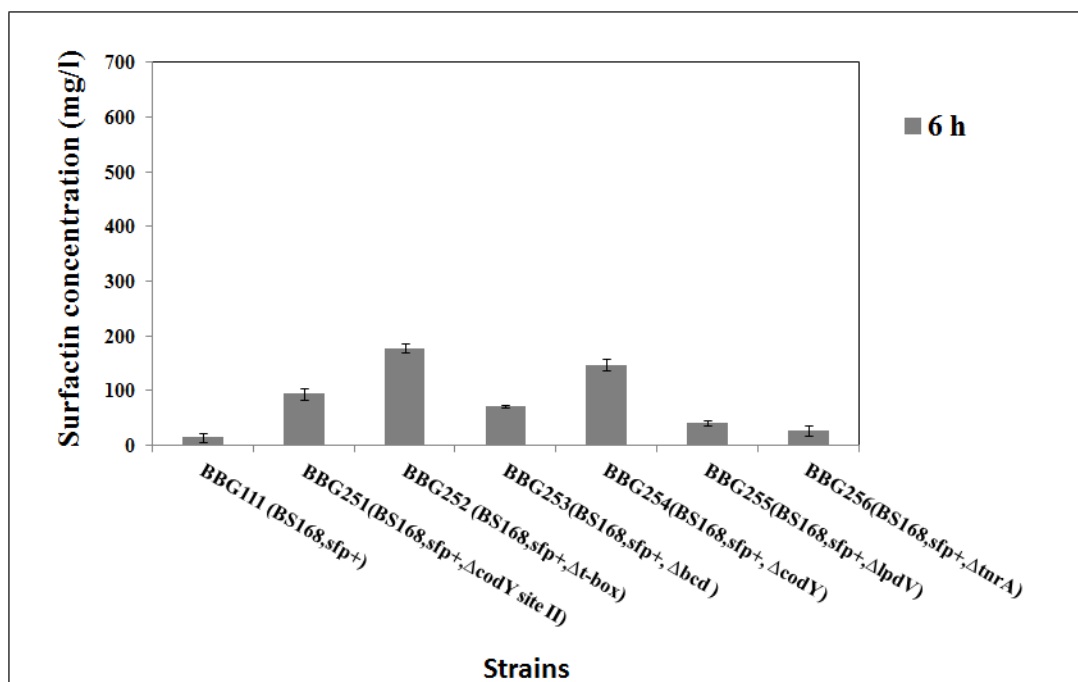


Figure 40 Surfactin production (mg/l) of BBG111 and its derivative strains in TSS medium supplemented with 16 amino acids at 37°C in Biolector® with shaking frequency of 1100 rpm. Results are expressed as mean value with standard deviation.

The surfactin concentration (mg/l) produced was normalized by the dryweight (g/l) of each strain at 6 hour to determine the specific surfactin yield by each strain. The specific surfactin yield of BBG111 was found to be 25±13 mg of surfactin per g of dry weight (DW) of cell biomass, whereas for BBG251, BBG252, BBG253, BBG254 and BBG256 was found to be 133±15, 235±7, 148±3, 529±80, 42±12 mg of surfactin per g of DW of cell biomass (Figure 41).

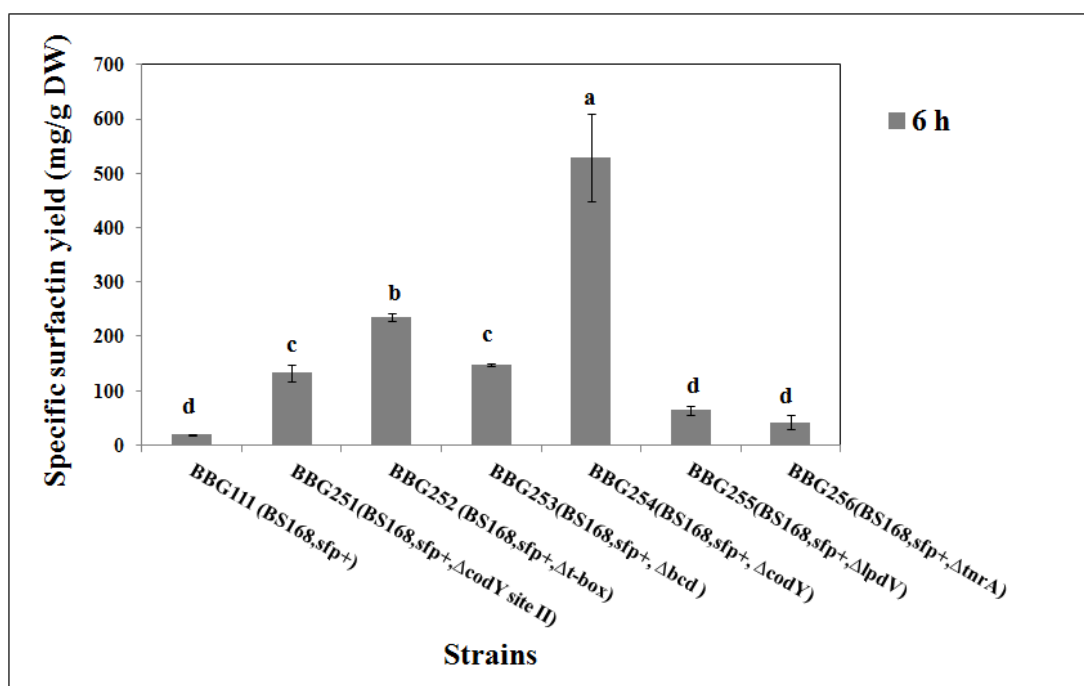


Figure 41 Specific surfactin yield (mg/g DW) in TSS medium supplemented with 16 amino acids with BGG111 and its derivative strains. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey’s multiple comparisons test.

The maximum surfactin yield was obtained with BGG254 (BS168, *sfp*⁺, Δ*codY*), which is 21-fold higher in comparison to control, BGG111. In Figure 40, BGG251 produced the maximum in comparison to other mutant strains but upon normalization the surfactin yield is found to be 9 times higher than control. The strain BGG256 showed the minimum enhancement in comparison to control. Through statistical analysis, it is determined that BGG254, BGG252, BGG251 and BGG253 have significantly enhanced the surfactin yield in comparison to BGG111. The low yield with BGG256 (BS168, *sfp*⁺, Δ*tnrA*) is due to dual repression, firstly CodY repressing both *ilv-leu* and *srfA* operon and secondly in the absence of ammonium is affected.

It was reported that in the presence of amino acids in the medium CodY can repress *srfA* operon (Serror and Sonenshein, 1996), which can justify the decline in surfactin yield in all mutants except BGG251 and BGG254. Only one site (site II) for CodY binding was deleted in BGG251 but the *P_{Ilv-Leu}* promoter contains four binding sites for CodY (Shivers and Sonenshein, 2004), in the promoter-proximal CodY-I+II (nt -84/-32) site that is a

high-affinity binding site of CodY, while the promoter-distal CodY-III (nt-154/-107) and CodY-IV (185/-168) sites are low-affinity ones can confirm the yield difference between BBG251 and BBG254.

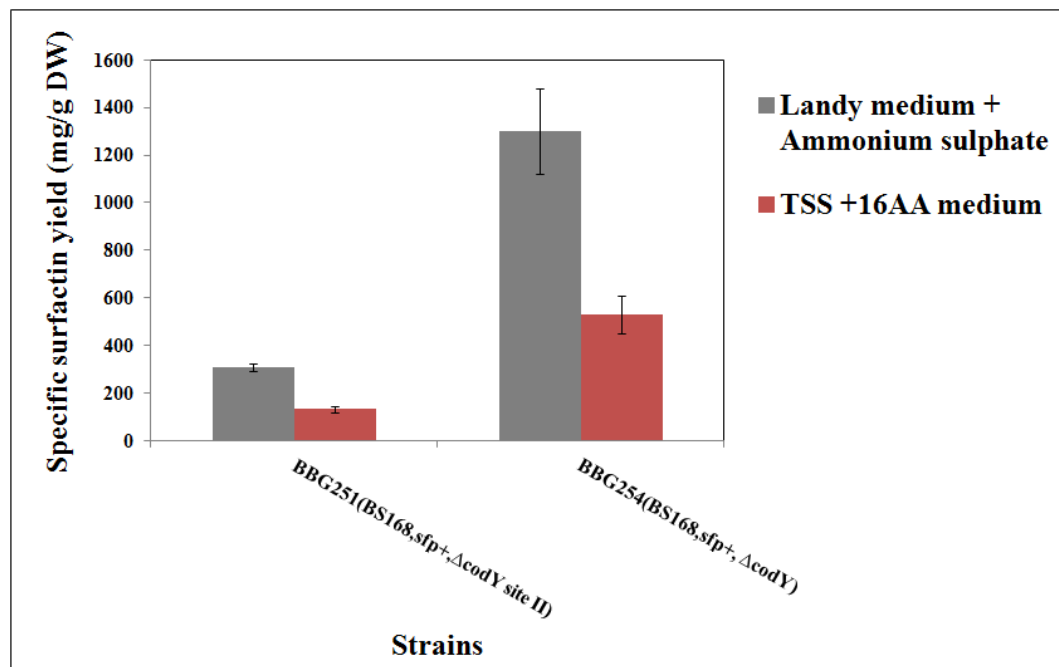


Figure 42 Comparative analysis on the basis of specific surfactin yield of BBG251 and BBG254 in Landy medium with ammonium sulphate and TSS medium supplemented with 16 amino acids. Results are expressed as mean value with standard deviation.

Comparative analysis of the specific surfactin yield, revealed that the yield obtained with BBG251 and BBG254 was 2.3-fold and 2.4-fold higher in Landy medium with ammonium sulphate in comparison to TSS supplemented with 16 amino acids (Figure 42). So, it can be suggested that besides CodY regulation, nutrient composition also plays an important role in enhancing the specific surfactin yield. This impact may be directly related to glucose concentration as its concentration is 5 times more in former medium in comparison to latter medium.

The regulation of the *ilv-leu* operon is dependent on the availability and nature of carbon and nitrogen sources in the medium, since these are involved in various global regulators i.e. CcpA (carbon), TnrA (nitrogen) and CodY (amino acids and GTP) and other regulators, which depends upon the influence of these global regulators and change in concentration of various intracellular metabolites (Brinsmade et al., 2014). Therefore, the

results also depend on the medium that is chosen. To analyze the specific surfactin production the strains were cultivated in two different media, which contain a different composition in carbon and nitrogen sources. Indeed, TSS medium with 16 amino acid is low concentrated in glucose (5 g/l), but very rich in organic nitrogen (4.5 g/l), unlike the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ is very rich in glucose (20 g/l) and mainly contains inorganic nitrogen (3.6 g/l) and a small amount of organic nitrogen (1 g/l yeast extract). As shown in Figure 42, the differences in medium composition lead to differences in specific surfactin yield. The production is increased for the mutants in glucose rich medium (Landy). Brinsmade et al found that *ilvB* transcript level was enhanced 80 times, 28 times and 19 times, when *codY*, *codY* binding site II and Leucine T-box deletion strain were grown in TSS medium supplemented with 16 amino acids (Brinsmade et al., 2010). These findings confirm the prediction made without knockout where an increase in pyruvate input leads to an increase of leucine production. Pyruvate is a direct result of glycolysis, an increase in glucose concentration in the medium may result in an increase of pyruvate intracellular concentration. Otherwise, the difference in nitrogen supply reveals a very different behaviour of the mutants.

3.2 Impact of oxygen transfer

Surfactin production highly depends on aeration and agitation. Surfactants at high concentration can affect oxygen transfer rate during fermentation (Hbid et al., 1996) and it was shown by Fahim et al., (2012) that an increase in volumetric oxygen transfer coefficient (k_{La}) can lead to an enhancement of surfactin production. However in this study whether impact of k_{La} on strains after genetic modification remains directly proportional was analyzed. In this study, two strains BBG258 (BSB1, sfp^+) and BBG260 (BSB1, sfp^+ , $\Delta codY$) were grown in Landy medium supplemented with ammonium sulphate. The strains were grown in 1000 ml flask with different filling ratio, V_L (0.05, 0.1 and 0.2) and after each time point (6 h and 10 h) the growth and surfactin production were analyzed. Suicidal flask strategy i.e. each flask was dedicated to each time point so that there is no alteration of volume after each time point. Experiments were carried out in triplicate and two biological replicates were performed. Results are expressed as mean value with standard deviation.

The volumetric oxygen transfer co-efficient in Biolector[®] and in shaking flask condition with each filling ratio was calculated with the equation 1 and 2 (Oxygenation page no. 48) The specific growth rate of both strains were calculated in different filling ratio in shake flask condition and in Biolector[®] have been shown in Table 14.

Table 14 Specific growth rate of BBG258 and BBG260 in different aeration conditions

REACTOR	V_L (ml/ml)	k_{La} (h^{-1})	SPECIFIC GROWTH RATE (h^{-1})	
			BBG258 (BSB1, sfp^+ : cm^2 , neo^+)	BBG260 (BSB1, sfp^+ : cm^2 , $\Delta codY$, neo^+)
Shake flask	0.05	632	0.55±0.04	0.52±0.02
	0.1	355	0.47±0.01	0.47±0.02
	0.2	199	0.45±0.06	0.48±0.01
Biolector	0.3	265	0.42 ± 0.01	0.46 ± 0.01

Volumetric oxygen transfer coefficient : k_{La}
Filling volume ratio : V_L

The specific growth rate of both strains varied with different filling ratio conditions. The growth rate increased in both strains with the decrease in V_L . This was well deduced from equation 1, according to which k_{La} will increase with decrease in V_L in shake flask

condition. The different conditions did not affect the specific growth rate of BBG260 in different conditions in comparison to BBG258 due to deletion of *codY*. The maximum specific surfactin yield obtained among two time points was plotted against different filling ratios. The maximum value was considered to minimize the effect of genetic modification on surfactin production. The maximum specific surfactin yield for BBG258 was found to be 495 ± 70 , 421 ± 23 and 234 ± 12 mg of surfactin per g DW of cell biomass whereas for BBG260 it was found to be 911 ± 4 , 1475 ± 196 and 1041 ± 48 mg of surfactin per g DW of cell biomass at filling volume 0.05, 0.1 and 0.2, respectively. It can be deduced that with BBG258 the yield declines with the increase in filling volume. The relationship between filling ratio and yield is not linear in the case of BBG260 as the yield was maximum with 0.1 but it declines with 0.2 and again increased drastically with 0.3. It can be inferred that the volumetric oxygen transfer coefficient might have a role on the yield of BBG260 as the geometry of the flask and the well of biolector was not same. The maximum specific surfactin yield for BBG258 was found to be 309 ± 48 , 268 ± 75 , 421 ± 23 and 495 ± 70 mg of surfactin per g DW of cell biomass whereas for BBG260 it was found to be 1041 ± 48 , 1277 ± 10 , 1475 ± 196 and 911 ± 4 mg of surfactin per g DW of cell biomass at k_{La} values of 199, 265, 355 and 632 h^{-1} respectively (Figure 43).

It can be inferred that with increase in k_{La} , the yield also increased, but until certain limit of k_{La} . It was found for BBG258 at a very high k_{La} value of 632 h^{-1} , the yield did not increase. While with BBG260 at 632 h^{-1} k_{La} , have negative influence on the surfactin yield. The biomass of both the strains (BBG258 and BBG260) are same at k_{La} values 355 and 632 h^{-1} , so there might be another factor regulating the specific surfactin yield of BBG260.

So, the maximum surfactin yield, was obtained with filling volume of 0.1 and at 250 rpm with this filling volume the volumetric oxygen transfer coefficient (k_{La}) will have positive impact on surfactin yield so these parameters were used for all the experiment in shake flask condition.

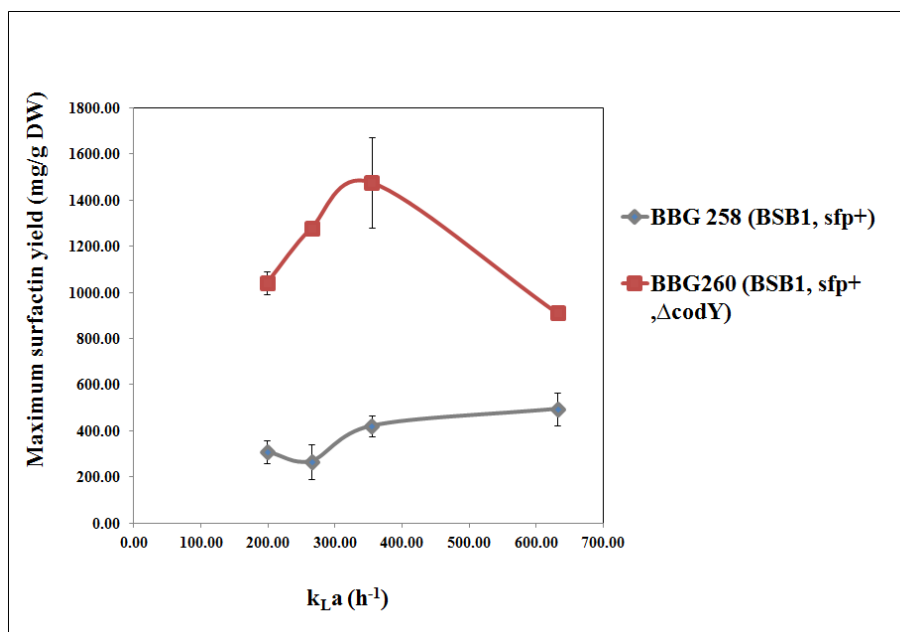


Figure 43 The role of volumetric oxygen transfer co-efficient ($k_L a$) in determining the maximum specific surfactin yield of BBG258 and BBG260. Results are obtained in shake flask and in Biolector[®] at 37⁰C and using Landy medium

The volumetric oxygen transfer coefficient is the key parameter controlling surfactin production and the selectivity of the parameters determining the aeration. Therefore, to scale-up fermentation process from flasks to large-scale fermentors, this parameter should be considered. Surfactin yield is clearly favoured at good oxygenation of the cells but there is an upper limit until which $k_L a$ has linear relationship with specific surfactin yield after that the relationship depends on the genetic construction of the strain.

3.3 Other factors

Besides glucose concentration, as shown in previous section that concentration of other nutrients can also have an impact on surfactin concentration. It was found that among the other nutrients (Mg^{2+} , K^+ , Mn^{2+} and Fe^{2+}), the concentration of Mg^{2+} influences the surfactin production the most (Wei et al., 2007).

The strains (Table 15) were grown in Landy medium with phosphate buffer (APPENDIX II) and at 6 h and 10 h the samples were analyzed for growth and surfactin analysis. The specific growth rate was measured during the exponential growth phase.

Table 15 BBG258 and its derivative strains used in this study

STRAIN	GENOTYPE	SOURCE
BBG258	BSB1, <i>sfp</i> ⁺ :cm ^f , neo ^f	This work
BBG259	BSB1, <i>sfp</i> ⁺ :cm ^f , Δbcd , neo ^f	This work
BBG260	BSB1, <i>sfp</i> ⁺ :cm ^f , $\Delta codY$, neo ^f	This work
BBG261	BSB1, <i>sfp</i> ⁺ :cm ^f , Δpdv , neo ^f	This work
BBG262	BSB1, <i>sfp</i> ⁺ :cm ^f , ΔtrA , neo ^f	This work

The specific growth rate of BBG258, BBG259, BBG260, BBG261 and BBG262 in this medium was found to be 0.37 ± 0.01 , 0.492 ± 0.007 , 0.481 ± 0.011 , 0.38 ± 0.008 and 0.598 ± 0.022 , respectively (Table 16).

Table 16 Specific growth rate of BBG258 and its derivative strains in different media

STRAINS	SPECIFIC GROWTH RATE (h ⁻¹)	
	LANDY MEDIUM WITH MOPS	LANDY MEDIUM WITH K-BUFFER
BBG258 (BSB1, <i>sfp</i> ⁺ :cm ^f , neo ^f)	0.42±0.01	0.37±0.01
BBG259 (BSB1, <i>sfp</i> ⁺ :cm ^f , Δbcd , neo ^f)	0.5 ± 0.02	0.492±0.007
BBG260 (BSB1, <i>sfp</i> ⁺ :cm ^f , $\Delta codY$, neo ^f)	0.46 ± 0.01	0.481±0.011
BBG261 (BSB1, <i>sfp</i> ⁺ :cm ^f , Δpdv , neo ^f)	0.39±0.03	0.38±0.008
BBG262 (BSB1, <i>sfp</i> ⁺ :cm ^f , ΔtrA , neo ^f)	0.5 ± 0.02	0.598±0.022

Comparative growth analysis revealed that the specific growth rate of BBG258 and BBG259 was reduced in this medium in comparison to its growth rate in Landy medium with MOPS, but contrary results were obtained with BBG262. The growth rate of the other mutants did not decrease in this medium in comparison to their growth rate in Landy-MOPS medium. Thus, it can be inferred that the decline in magnesium concentration might have affected the growth rate of BBG258 and BBG259.

The surfactin concentration (mg/l) for BBG258, BBG259, BBG260, BBG261 and BBG262 was found to be 68 ± 5 , 392 ± 35 , 1243 ± 153 , 109 ± 31 and 542 ± 91 at 6 hour ; at 10 hour the surfactin concentration was found to be 221 ± 83 , 611 ± 115 , 2289 ± 318 , 252 ± 21 and 1248 ± 177 and at 24 hour the surfactin concentration was found to be 694 ± 93 , 736 ± 103 , 2911 ± 141 , 627 ± 74 and 1639 ± 358 respectively (Figure 44). The maximum concentration was produced by BBG260 (BSB1, *sfp*⁺, Δ *codY*) and it was found to be 18 times; 10 times and 4 times more at 6 hours; 10 hours and 24 hours, respectively in comparison to BBG258.

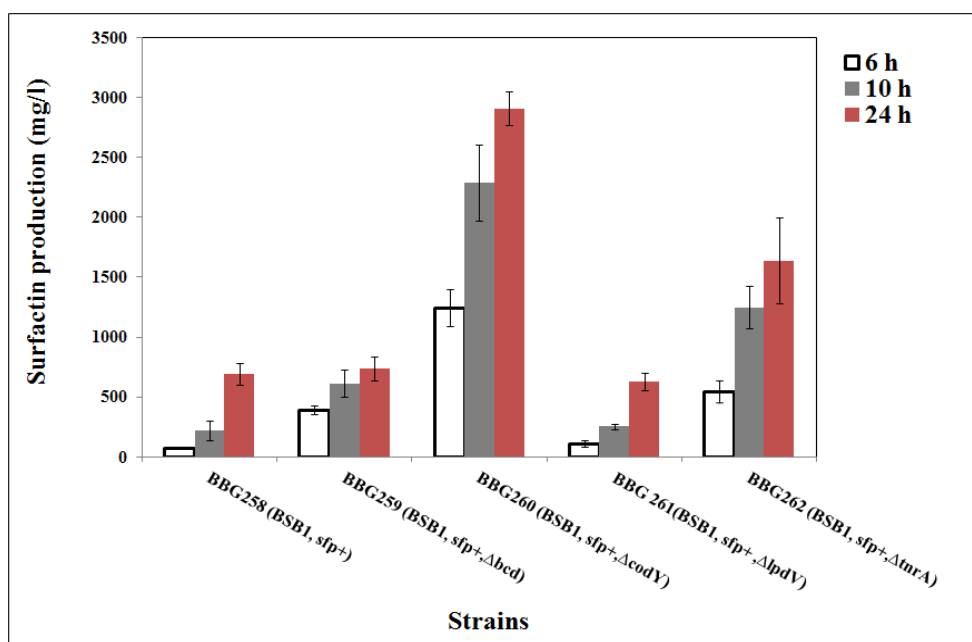


Figure 44 Surfactin production (mg/l) by BBG258 and its derivative strains in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation.

The surfactin concentration was normalized with dry weight of the strains at each time point and the specific surfactin yield was determined at each time point. This value was represented in mg of surfactin per g dry weight of cell biomass.

The specific surfactin yield (mg of surfactin/g DW) for BBG258 , BBG259, BBG260, BBG261 and BBG262 was found to be 257±11, 414±55, 1483±233, 368±96 and 276±32 at 6 hours; at 10 hours the surfactin yield was found to be 526±80, 360±69, 1556±123, 782±51 and 560±70 and 24 hours the surfactin yield was found to be 1480±195, 809±114, 3138±112, 981±112 and 2526±388, respectively (Figure 45). The maximum yield was obtained with BBG260 (BSB1, *sfp*⁺, Δ *codY*) and the yield was enhanced 6 times, 3 times and 2 times at 6 hours, 10 hours and 24 hours in comparison to control (BBG258).

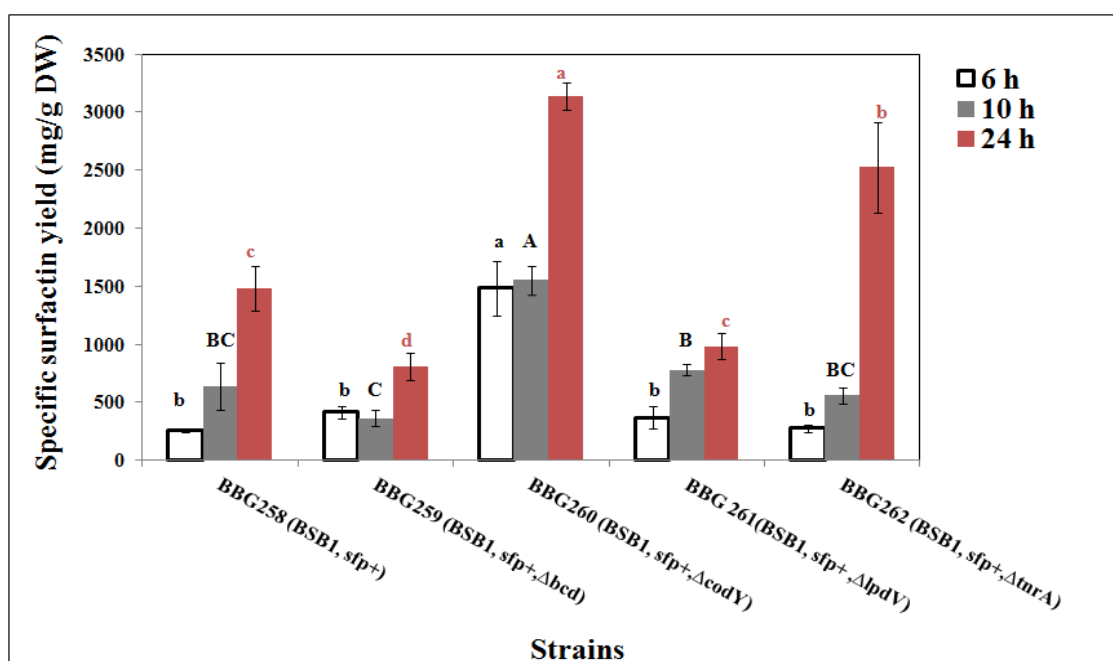


Figure 45 Specific surfactin yield obtained with BBG258 and its derivative strains in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey’s multiple comparisons test.

Statistical analysis revealed that BBG260 at both 6, 10 and 24 hours and BBG262 at 24 hour had a significant impact on the enhancement of specific surfactin yield in comparison to BBG258.

It can be inferred from surfactin analysis, although Sfp protein needs magnesium ion as a cofactor (Wei et al., 2007), magnesium ion did not affect the production nor the yield of surfactin with BBG260 in comparison to BBG258, whereas the yield of BBG259 and BBG262 was affected as the end of 10 hours the yield was less or almost equal to one of BBG258 (Figure 37) (Figure 39).

The native promoter of *srfA* operon was exchanged with a constitutive promoter and affect of this new promoter was studied in this medium. The study was carried out to analyze whether the new promoter can enhance further the production of BBG260 in this new medium.

To study this objective two new strains (Table 17) were constructed with transformation of BBG258 and BBG260 with pBG401 (Figure 28) to replace the native promoter (P_{srfA}) by a constitutive promoter (P_{repU}). This constitutive promoter regulates the replication gene *repU* in *S. aureus* (Coutte et al., 2010).

Table 17 Genotype of BBG276 and BBG278 strain with a constitutive promoter

STRAIN	GENOTYPE	SOURCE
BBG276	BBG258, P_{repU} -erm ^R ::HaeIII-srfA	This work
BBG278	BBG260, P_{repU} -erm ^R ::HaeIII-srfA	This work

The specific growth rate and surfactin production was analyzed for these two strains in Landy-phosphate buffer medium. The specific growth rate was measured during the exponential growth phase and surfactin concentration was measured at 6 hours and 10 hours, respectively.

Table 18 Specific growth rate of BBG276 and BBG278 in Landy-phosphate buffer medium

STRAINS	SPECIFIC GROWTH RATE (h ⁻¹)
BBG276 (BBG258, P _{repU} -erm ^R ::HaeIII- <i>surfA</i>)	0.397±0.002
BBG278 (BBG260, P _{repU} -erm ^R ::HaeIII- <i>surfA</i>)	0.472±0.016

The specific growth rate of BBG276 and BBG278 was found to be 0.397±0.002 and 0.472±0.016 h⁻¹ (Table 18). The specific growth rate of these new mutants did not vary in comparison to its parental strain. The replacement of the promoter did not affect the growth rate of the strain.

The surfactin concentration (mg/l) for BBG276 and BBG278 was found to be 116±33 and 731±139 at 6 hours, whereas at 10 hours the surfactin concentration was found to be 245±111 and 1435±179, respectively (Figure 46). The surfactin production of BBG278 was enhanced 6 times at both time points in comparison to control.

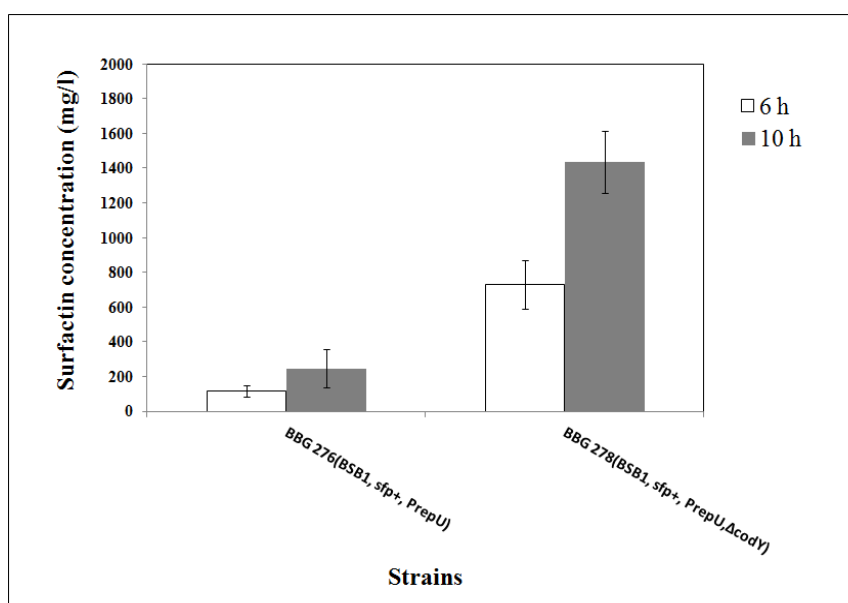


Figure 46 Surfactin production (mg/l) by BBG276 and BBG278 in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation.

The surfactin concentration was normalized with dry weight of the strains at each time point and the specific surfactin yield was determined at each time point. This value was represented in mg of surfactin per g dry weight of cell biomass.

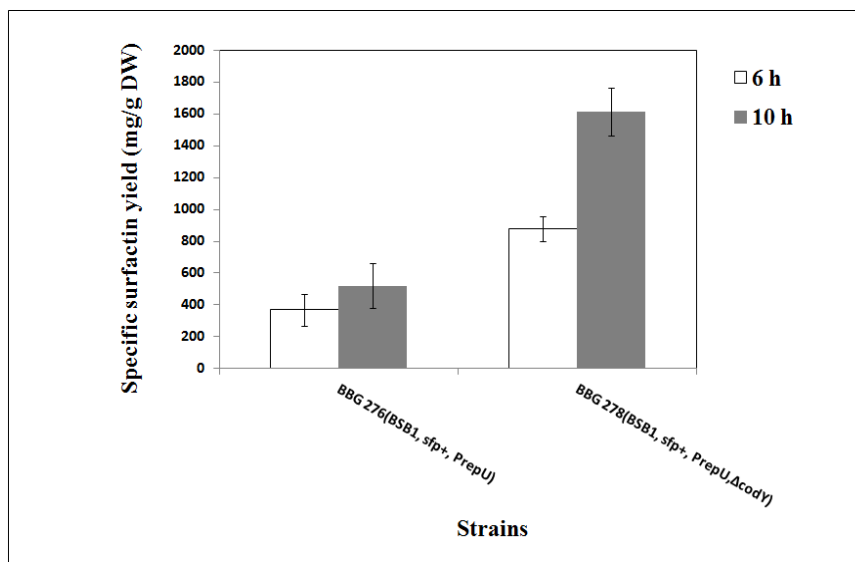


Figure 47 Specific surfactin yield obtained with BBG276 and BBG278 in Landy medium with phosphate buffer at 37°C in shake flask condition with 250 rpm. Results are expressed as mean value with standard deviation.

The specific surfactin yield of BBG276 and BBG278 was found to be 369 ± 101 and 878 ± 79 at 6 hours whereas at 10 hours the surfactin concentration was found to be 521 ± 142 and 1613 ± 150 respectively (Figure 47). The surfactin yield of BBG278 in comparison to control was enhanced 2 times and 3 times at both 6 hours and 10 hours respectively.

Major conclusions can be drawn from the above results. The promoter replacement did not have a great impact on the surfactin production. As BBG258 and BBG260 harboring the surfactin native promoter enhanced surfactin yield 1.4 times and 1.6 times, in comparison to its counterpart BBG276 and BBG278 having constitutive promoter at 6 hour (Figure 45). But at 10 hours the surfactin yield of BBG258 and BBG260 was found to be almost equal to BBG276 and BBG278, although the constitutive promoter (P_{repU}) gave 3 fold and 15 fold higher production of iturin A (K Tsuge et al., 2001) and mycosubtilin (Leclère et al., 2005) in comparison to corresponding strains with native promoter.

3.4 Publication and Valorization of these results

ORAL COMMUNICATION

Dhali.D, Coutte.F, Niehren.J, Versari.C, John.M, Bidenko.V, Auger.S, Jacques.P. **Metabolic engineering for the overproduction of surfactin: an integrative approach**, International Conference on Metabolic Engineering in Bacteria, Amsterdam, The Netherlands, 16-17 April 2015.

POSTER COMMUNICATION

Coutte, F., John, M., Dhali, D., Béchet, M., Niehren J., Jacques P. **Genetic engineering of the branched chain amino-acid biosynthetic pathway for the overproduction of surfactin by *Bacillus subtilis* : Application of biocomputing predictions**, 10th European Symposium on Biochemical Engineering Science, Lille, France, 8-10 September 2014.

4. ROLE OF PYRUVATE FLUX

Besides single knockout prediction from the reaction network (Figure 34), the prediction based on flux was also provided (Table 8). The first solution of this prediction suggested that an increase in pyruvate flux can lead to an enhancement of leucine. In the glycolytic pathway (Figure 15), there are various genes which are essential for the growth of *B. subtilis* (Commichau et al., 2013). Besides these genes four genes were selected within the glycolytic pathway to analyze their role in surfactin production. All the partners in the European project also agreed to work on similar set of genes to have common strains for future experiments.

4.1 Mutant constructed

The glycolytic pathway mutants were constructed in *B. subtilis* BSB1 strain by the group of Prof. J Stülke (University of Göttingen, Germany) (Table 19).

These genes include *pgcA*: responsible for the interconversion of glucose-6-phosphate to α -glucose-1-phosphate; *pfkA*: catalyzes the conversion of fructose-6-phosphate to fructose-1, 6-bisphosphate; *pyk*: converts phosphoenol pyruvate to pyruvate and *pgi*: is an isomerase responsible for reversible reaction between glucose-6-phosphate and fructose-6-phosphate.

Table 19 Deletion mutants obtained with *Bacillus subtilis* BSB1 strain with control based on glycolytic pathway

STRAIN	GENOTYPE	SOURCE
GP1743	BSB1, $\Delta pgcA:kan^r$	Prof. J. Stülke, University of Göttingen, Germany
GP1744	BSB1, $\Delta pfkA:kan^r$	
GP1745	BSB1, $\Delta pyk:kan^r$	
GP1746	BSB1, $\Delta pgi:kan^r$	

These mutants lack *sfp* gene and were thus transformed with pBG400 for the insertion of *sfp* gene to analyze the impact of the individual genes on surfactin production. The mutants were renamed after *sfp* gene insertion as shown in Table 20.

Table 20 Integration of *sfp* gene in the glycolytic mutants obtained with *Bacillus subtilis* BSB1 strain

STRAIN	GENOTYPE	SOURCE
BBG266	BSB1, <i>sfp</i> ⁺ : <i>cm</i> ^f , Δ <i>pfkA</i> : <i>kan</i> ^f	This work
BBG267	BSB1, <i>sfp</i> ⁺ : <i>cm</i> ^f , Δ <i>pgi</i> : <i>kan</i> ^f	This work
BBG269	BSB1, <i>sfp</i> ⁺ : <i>cm</i> ^f , Δ <i>pyk</i> : <i>kan</i> ^f	This work

4.2 Growth and surfactin analysis

The glycolytic mutants obtained with *B. subtilis* BSB1 strain shown in Table 14 were cultivated in high-throughput system of fermentation Biolector[®]. Cultures were performed in both classical Landy medium and Amber medium at 37°C with shaking frequency of 1100 rpm. During the exponential phase, the glycolytic genes are highly expressed in order to correlate their influence on surfactin production the samples were withdrawn at 6 h and 8 h of culture.

4.2.1 Growth analysis

Growth kinetics study of the mutant strains were carried out during the exponential growth phase in classical Landy medium and Amber medium using the BioLecton V.2.4.1.0 software (m2p-labs GmbH, Germany). The detailed composition of the media used has been provided in APPENDIX II. In both media, besides glucose, another carbon source has been used, as the mutants could not grow properly if only glucose is present. Classical Landy medium contains glucose and glutamic acid whereas AMBER medium contains glucose, L-malate and glutamic acid. Previous studies have revealed that malate is co-

metabolized with glucose, and malate metabolism is not prone to glucose dependent carbon catabolite repression (Buescher et al., 2012).

When BBG258 was grown in classical Landy medium, the specific growth rate was found to be $0.489 \pm 0.002 \text{ h}^{-1}$. But there was a decline in the specific growth rate of the mutants and it was found to be 0.232 ± 0.014 ; 0.406 ± 0.017 ; 0.394 ± 0.006 and $0.335 \pm 0.023 \text{ h}^{-1}$ for BBG266; BBG267; BBG268 and BBG269. It was found that there was a significant decline in specific growth rate of BBG266 where the gene *pfkA* was deleted.

When BBG258 was grown in Amber medium, the specific growth rate was found to be $0.355 \pm 0.002 \text{ h}^{-1}$. But there was a decline in the specific growth rate of the mutants and it was found to be 0.312 ± 0.013 ; 0.394 ± 0.013 ; 0.319 ± 0.003 and $0.37 \pm 0.002 \text{ h}^{-1}$ for BBG266; BBG267; BBG268 and BBG269.

An overview of specific growth rate of BBG258 and its derivatives in different media have been provided in Table 21. The deletion of the respective genes produces viable cells as the genes were not essential for the viability of the cells (Commichau et al., 2013).

Table 21 Specific growth rate of glycolytic mutant strains and BBG258 in Landy medium and AMBER medium

STRAINS	SPECIFIC GROWTH RATE (h^{-1})	
	LANDY MEDIUM	AMBER MEDIUM
BBG258 (BSB1, <i>sfp</i> ⁺ :cm ^r , neo ^r)	0.489±0.002	0.355±0.002
BBG266 (BSB1, <i>sfp</i> ⁺ :cm ^r , Δ <i>pfkA</i> :kan ^r)	0.232±0.014	0.312±0.013
BBG267 (BSB1, <i>sfp</i> ⁺ :cm ^r , Δ <i>pgi</i> :kan ^r)	0.406±0.017	0.394±0.013
BBG268 (BSB1, <i>sfp</i> ⁺ :cm ^r , Δ <i>pgcA</i> :kan ^r)	0.394±0.006	0.319±0.003
BBG269 (BSB1, <i>sfp</i> ⁺ :cm ^r , Δ <i>pyk</i> :kan ^r)	0.335±0.023	0.37±0.002

These genes are involved in the regulation of the glycolytic pathway and determine the flux towards pyruvate. Beside their role in the glycolytic pathway they can indirectly regulate various other pathways in *B. subtilis*. Growth analysis of BBG258 showed that the specific growth rate of this strain is better in Landy medium in comparison to AMBER medium. The reason for such a behaviour may be due to fact that glucose concentration

(20 g/l) in Landy medium is 5 times more than glucose concentration (4 g/l) in Amber medium irrespective to the presence of other carbon sources and may be also due to presence of yeast extract in Landy medium.

The gene *pfkA* is responsible for the synthesis of fructose-1, 6-bisphosphate. There was a marked reduction in specific growth rate of this mutant in Landy medium than in Amber medium with respect to their control. In both media, the central pathway for the utilization is blocked due to deletion of respective genes but in Amber due to presence of malate which is responsible for gluconeogenesis, it can counter-balance the loss of glucose. The disruption of *pgi* did not affect the specific growth rate of the strain in both media as metabolic flux can flow through alternative pathway (pentose phosphate pathway) for the synthesis of respective metabolites (Figure 52). The deletion of *pgcA* also has no effect on the specific growth rate of the strain in both media as it is responsible for the conversion of glucose-1-phosphate to glucose-6-phosphate and did not affect the metabolic flow in the central pathway. In Landy medium, the strain with deletion of *pykA* has reduced specific growth rate but in AMBER medium it has no effect.

4.2.2 Surfactin analysis

The impact of gene deletion on surfactin production was analyzed. The surfactin concentration at each time point (6 h and 8 h) was determined through RP-HPLC.

It can be observed from Figure 48, that all the mutant strains produced surfactin with varied concentration in classical Landy medium. The control strain (BBG258) produced 746 ± 29 mg/l and 1082 ± 50 mg/l of surfactin at 6 hours and 8 hours, respectively. At both time points the concentration of surfactin (mg/l) by the glycolytic mutant strains was less in comparison to BBG258. The maximum surfactin production among the mutants was obtained with BBG268 (*BSB1*, *sfp*⁺, Δ *pgcA*) at both time points with a concentration of 483 ± 7 mg/l and 1018 ± 78 mg/l. The figure does not represent the normalization concentration with dry weight of specific strain; this representation has been provided in the later section.

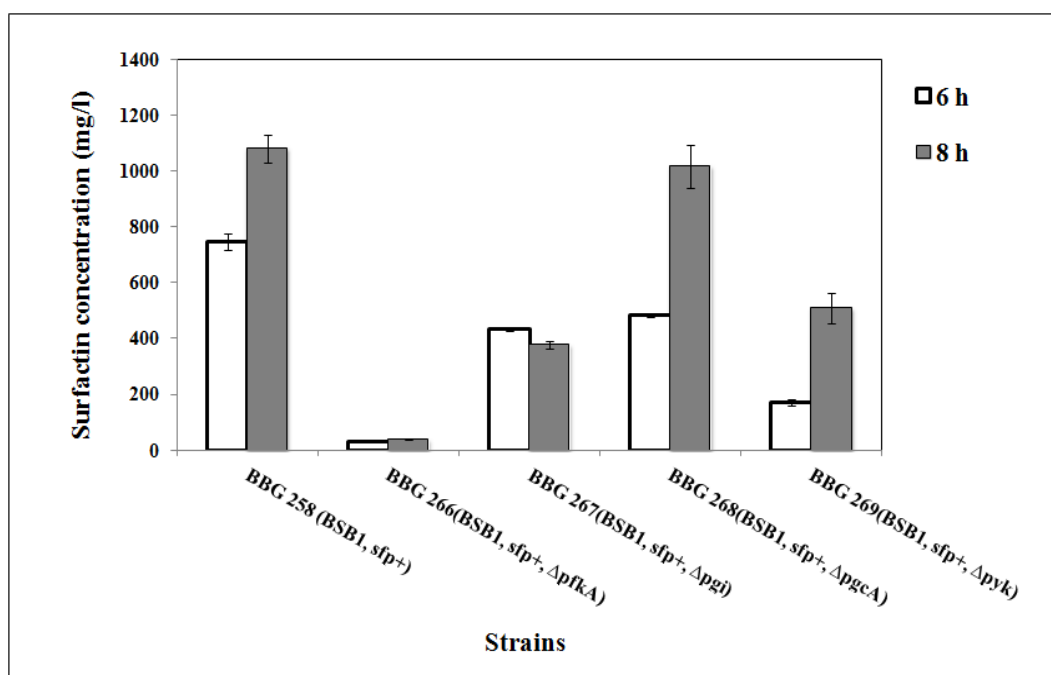


Figure 48 Surfactin production (mg/l) of BBG258 and its derivative strains in classical Landy medium at 6 hours and 8 hours, respectively in Biolector[®] at 37°C with shaking frequency 1100 rpm. Results are expressed as mean value with standard deviation.

When the strains were analyzed in AMBER medium similar variation in surfactin concentration was produced as shown in Figure 49. At 6 hours and 8 hours, BBG258 produced surfactin at a concentration of 456±20 mg/l and 746±20 mg/l. At both time points the glycolytic mutant strains produced less surfactin in comparison to BBG258. The maximum surfactin production among the mutants was obtained with BBG269 (*BSB1, sfp⁺, ΔpykA*) at 6 hours with a concentration of 289±6 mg/l but at 8 hour both BBG268 (*BSB1, sfp⁺, ΔpgcA*) and BBG269 (*BSB1, sfp⁺, ΔpykA*) produced equivalent amount of surfactin with a concentration of 547±21 mg/l and 545±21 mg/l, respectively. Both figures does not represent the normalization with dry weight of a specific strain.

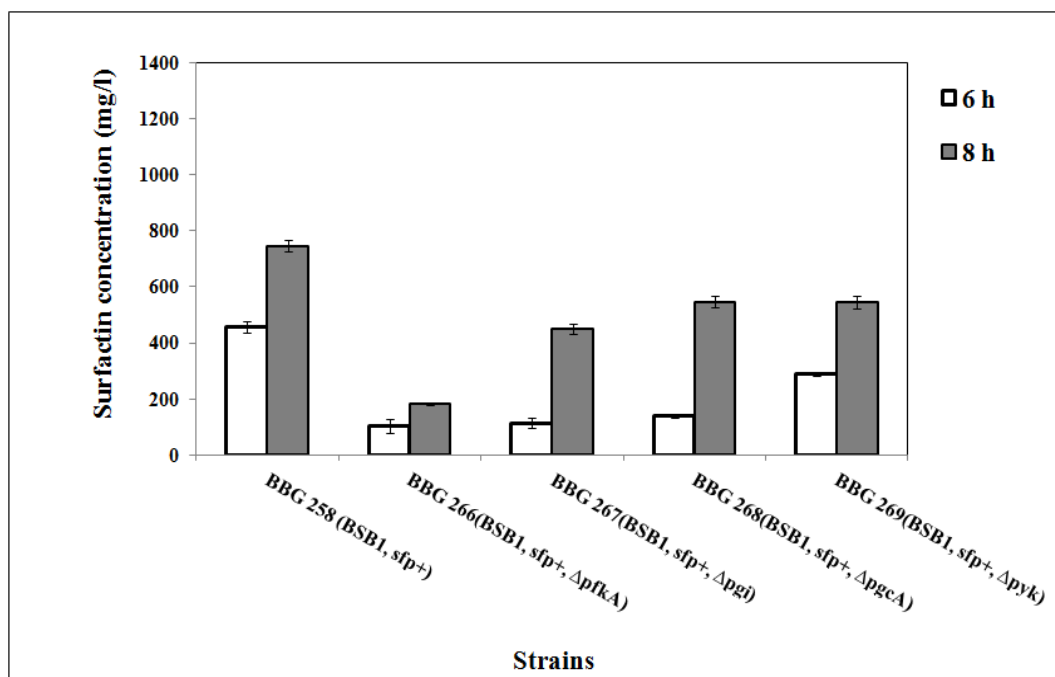


Figure 49 Surfactin production (mg/l) of BBG258 and its derivative strains in AMBER medium at 6 hours and 8 hours, respectively in Biolector[®] at 37°C with shaking frequency 1100 rpm. Results are expressed as mean value with standard deviation.

Thus, it can be deduced from Figure 48 and Figure 49 that modifying glycolytic pathway has a definitive effect on the surfactin production. It can be observed that the surfactin production of BBG258 in classical Landy medium was more than in AMBER medium, it is due to variation in glucose concentration. But the mutant strain BBG266 (*BSB1*, *sfp*⁺, *ΔpfkA*) which is the least producer among the strains in both media was found to be better producer in AMBER medium than in classical Landy medium due to the presence of malate in the nutrient composition in the former medium which led to gluconeogenesis resulting in synthesis of few metabolites which could not be synthesized by glycolysis due to deletion of respective gene (Figure 15).

The surfactin production was normalized with dry weight (DW) of each strain. The normalized surfactin production was renamed as surfactin yield (mg of surfactin per gm DW).

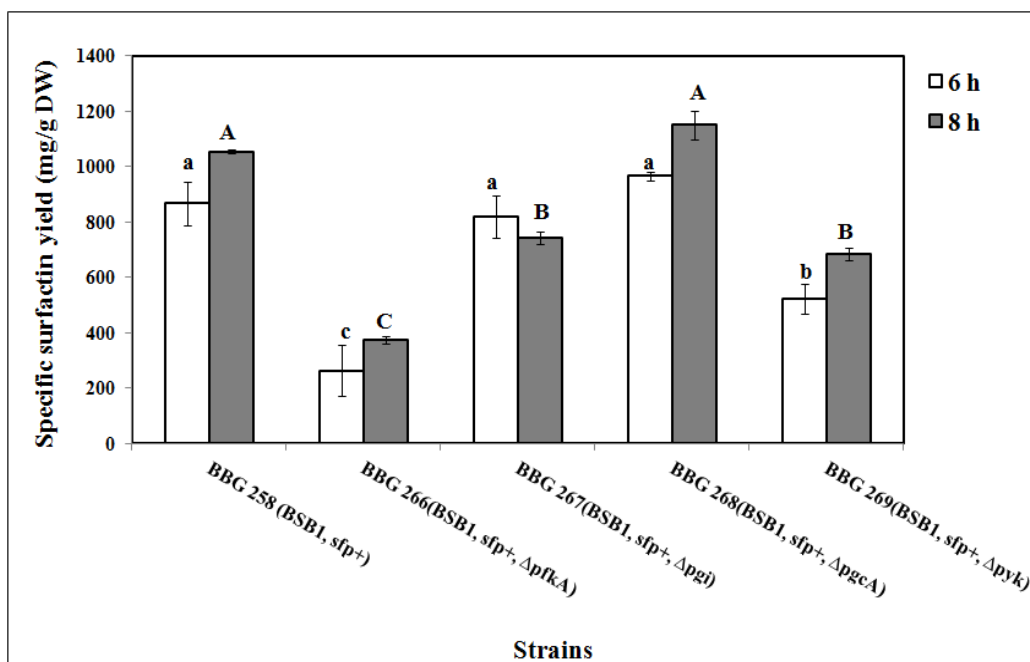


Figure 50 Specific surfactin yield obtained with BBG258 and its derivative strains in classical Landy medium at 6 hour and 8 hour respectively. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test.

The specific surfactin yield of BBG258 was found to be 868 ± 78 and 1054 ± 7 mg of surfactin per gm of dry weight (DW) of cell biomass at 6 hours and 8 hours, respectively in classical Landy medium. For BBG266, BBG267, BBG268, and BBG269 specific surfactin yield was found to be 264 ± 94 ; 817 ± 76 ; 966 ± 15 ; 523 ± 54 mg of surfactin per gm of DW of cell biomass at 6 hours while at 8 hours the specific surfactin yield was 375 ± 15 ; 744 ± 23 ; 1150 ± 51 ; 684 ± 22 for the respective strains (Figure 50). The minimum surfactin yield was obtained with BBG266 (*BSB1, sfp⁺, ΔpfkA*), which is 3 fold lower in both time points in comparison to control, BBG258. Although in Figure 48, the strain BBG268 (*BSB1, sfp⁺, ΔpgcA*) did not produce equivalent amounts of surfactin concentration at 6 hours in comparison to control but upon normalization the surfactin yield of this strain is equal to the surfactin yield of BBG258 in classical Landy medium.

Glutamic acid can positively influence the surfactin yield and it can be well correlated from Figure 39 and Figure 50. The surfactin yield was increased 3.5 folds in Landy medium with glutamic acid in comparison to Landy medium with ammonium sulphate.

When similar experiment was performed in AMBER medium with BBG258 and its derivative strains, the specific surfactin yield of BBG258 was found to be 760 ± 34 and 941 ± 31 mg of surfactin per gm of cell DW at 6 hour and 8 hour of incubation. The surfactin yield for BBG266, BBG267, BBG268 and BBG269 was found to be 341 ± 79 ; 399 ± 72 ; 711 ± 18 and 369 ± 75 mg of surfactin per gm of cell dry weight respectively at 6 hour while at 8 hour the specific surfactin yield was found to be 325 ± 26 ; 421 ± 12 ; 913 ± 36 and 606 ± 24 mg of surfactin per g DW for the respective strains (Figure 51).

The surfactin yield obtained with the strain BBG266 (BSB1, *sfp*⁺, $\Delta pfkA$) was lowest among the glycolytic mutants. The specific surfactin yield of BBG266 was found to be 2 times and 3 times lower than the control strain, BBG258 at 6 hour and 8 hour, respectively. The surfactin yield of strain BBG268 (BSB1, *sfp*⁺, $\Delta pgcA$) was equivalent to BBG258 at both time points. In Figure 49, the surfactin concentration obtained with BBG268 was lower than BBG258 in both time points. This observation was due to the fact that the growth of BBG268 was lower than BBG258 in both the time points, which reflected in equal yield of surfactin by both strains.

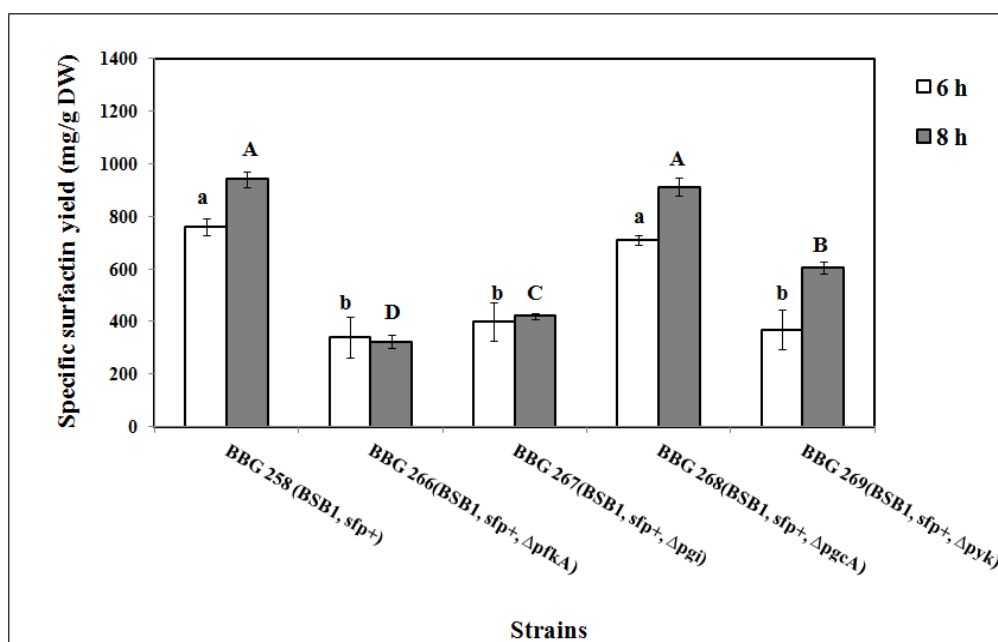


Figure 51 Specific surfactin yield obtained with BBG258 and its derivative strains in AMBER medium at 6 hours and 8 hours, respectively. Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with Tukey's multiple comparisons test.

Statistical analysis of Figure 50 and Figure 51, revealed that at 6 hour there is a significant decline in surfactin yield with BBG266 and BBG269 in Landy medium but in AMBER medium there is a significant decrease in yield with the strains BBG266, BBG267 and BBG269, respectively in comparison to BBG258. At 8 hour, there is a significant decrease with BBG266, BBG267 and BBG269, respectively, in both the condition in comparison to BBG258.

Thus, the glycolytic mutants have negative effect on surfactin production in both the media. Although the yield of surfactin was obtained more in classical Landy medium in comparison to AMBER medium with all mutant strains along with control. This observation reflected the fact that high glucose concentration can lead to high surfactin production, in this study which was present in classical Landy medium.

4.3 Carbon metabolism and surfactin production

Glycolytic pathway is responsible for the uptake of carbon source and is simultaneously distributed to various pathways. An important starting point for metabolic engineering is to determine the flow of carbon flux towards the metabolic link of various pathways. Pyruvate forms this linkage, as carbon flux enters glycolytic pathways and is distributed to various pathways i.e. production of lactate; acetate; acetoin; ethanol; 2, 3-butandiol; branched chain amino acids etc (Romero et al., 2007). Different studies have been carried out to determine the role of glycolytic pathway in the production of various secondary metabolites. In *E. coli*, the production of lycopene was limited by the concentration of glyceraldehyde-3-phosphate. This problem was overcome by reducing the metabolic flux to pyruvate by the deletion of *pykA*, responsible for the conversion of phosphoenol pyruvate to pyruvate (Farmer and Liao, 2001). Similar study was performed with *Streptomyces clavuligerus* to enhance the biosynthesis of clavulanic acid. Here, two genes *gap1* and *gap2* were deleted to increase the pool of glyceraldehyde-3-phosphate, which is one of the precursors for the production clavulanic acid (Olano et al., 2008). Thus, the primary metabolism forms a determining factor for the supply of precursors required for production of secondary metabolites. Glucose is a good source of carbon as it can support growth as well as other physiological processes i.e. surfactin production. The regulation of surfactin production is very complex. The synthetase requires various monomers for the synthesis of this non-ribosomal lipopeptide. Previous studies suggested that surfactin

production depends on cell density and carbon concentration. Until now no data have been presented between surfactin production and central carbon metabolism. There are various genes within central carbon metabolism pathway which are essential for the growth of *B. subtilis*. These genes include *gapA*; *pgm* and *eno*, which are responsible for the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate; 3-phosphoglycerate; 2-phosphoglycerate to phosphoenolpyruvate (Commichau and Stülke, 2008) (Figure 52).

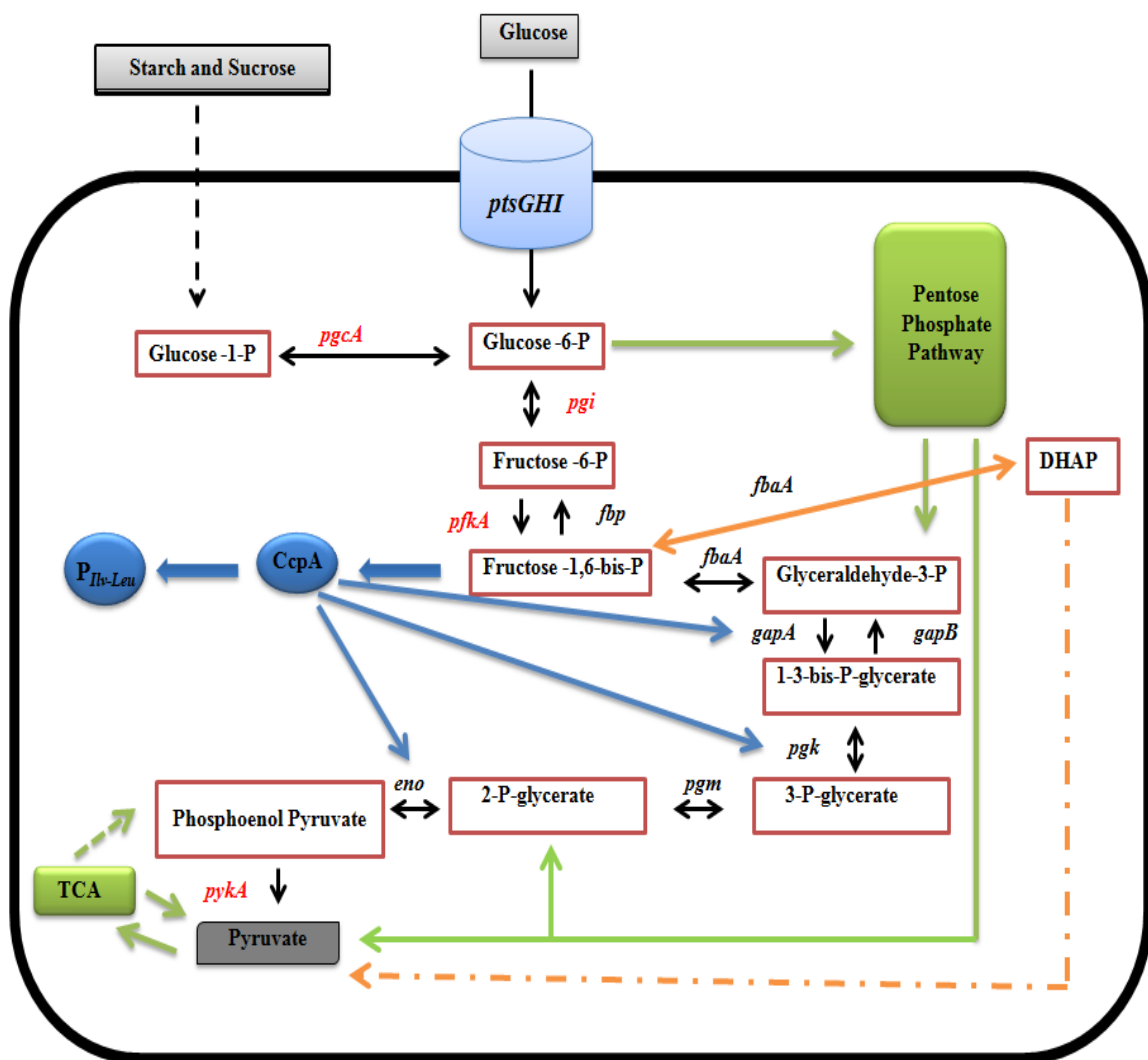


Figure 52 Regulation of glycolytic pathway and genes marked in red were deleted to analyze their roles in surfactin production

So it was not possible to study the effect of these genes on surfactin production. A particular gene set selected as shown in Table 19, showed diverse effects on surfactin

production. The production analysis showed that effect of nutrient composition can also have additional or counter balancing effect on the glycolytic mutants.

Based on the nutrient composition, the growth as well as the surfactin yield of BBG258 (control strain) was better in classical Landy medium in comparison to AMBER medium in all time points (6 hour and 8 hour). But the glycolytic mutants showed variation in specific growth rate in Landy medium but the differences in AMBER medium was closely clustered. The specific surfactin yield of these mutants in Landy medium was better than in AMBER medium at both time points reflecting that high flux of glucose is essential for the higher production of precursors required for surfactin biosynthesis.

Upon uptake of glucose by PtsGHI, glucose is phosphorylated to glucose-6-phosphate, although the biosynthesis of this molecule can also take place through the conversion of glucose-1-phosphate by *pgcA*. This gene does not lie within the central metabolic pathway and the substrate for this gene is provided by other carbon sources i.e. starch and sucrose (Figure 52). This gene is responsible for the conversion of glucose-1-phosphate to UDP-glucose, which is essential for the synthesis of lipoteichoic acid, major and minor teichoic acid. These components are the constituents for the cell envelope biosynthetic process. This gene produces a nucleotide sugar, which is a substrate for the *eps* pathway for the production of exopolysaccharide in *B. subtilis* and surfactin also plays a vital role in the arrangement of aerial structures on biofilm of *B. subtilis* (Vlamakis et al., 2013). But no relation has been established between this gene and surfactin biosynthesis. The results (Figure 50 and Figure 51) showed that the gene cannot lower the flux for the synthesis of precursors required for the synthesis of surfactin.

The gene *pgi* is responsible for a reversible reaction between glucose-6-phosphate and fructose-6-phosphate. Presence of glucose can increase the expression of *gap* gene and *pgk* operon but the expression decreased in the absence of glucose. But this glucose dependent activity was not observed for genes *pgi*, *pfk*, *fbaA* and *pykA* (Tobisch et al., 1999). It was reported that a mutant of *pgi* have poor growth in medium where carbon source is solely glucose as the flux is directed towards the pentose phosphate pathway (Lin and Prasad, 1974) (Figure 52). Similar behavior was observed in this study that when BBG267 (BSB1, *sfp*⁺: cm^r, Δ *pgi:kan*^r) was grown in classical Landy medium, there was a slight decrease in specific growth rate but no effect was observed when it was grown in AMBER medium. The reason for such a behavior is due to the fact that classical Landy has only glucose but AMBER media contains both glucose and malate. This gene is constitutively expressed

and an increase in intracellular concentration of this protein was observed upon salt stress (Kohlstedt et al., 2014a). But the salt stress condition was not imposed in this study. Thus, it can be inferred that an alteration of surfactin yield was observed with *pgi* mutant in comparison to control is solely due to decline in metabolic flux and less synthesis of precursors required for the biosynthesis of surfactin.

The conversion of fructose-6-phosphate to fructose-1, 6-bisphosphate is facilitated by *pfkA* whereas *fbp* is responsible for the reverse reaction. The former reaction is ATP-dependent that represents the initial step dedicated to glycolysis. Upon deletion of *pfkA* the doubling time was increased in comparison to control in both minimal and rich media (Muñoz-Márquez and Ponce-Rivas, 2010). This result is coherent with the observation in this study, as in both classical Landy and Amber media the strain BBG266 (BSB1, *sfp*⁺: cm^r, Δ *pfkA*: *kan*^r) showed growth defects. The expression of this gene is induced by the presence of glucose thus regulating the direction of flux (Holger Ludwig et al., 2001). Regulation of gene expression is controlled by messenger RNA turnover mechanism in bacteria. This mechanism is mediated by the role of various RNA binding proteins, noncoding RNAs and action of various ribonucleases (RNases) (Newman et al., 2012). The glycolytic enzymes enolase and phosphofructokinase (PfkA) along with four proteins (RNase Y, PNPase, RNase J1 and RNase J2) and RNA helicase CshA are involved in the metabolism of RNA and are known as degradasome (Lehnik-Habrink et al., 2010). This association forms a link between glycolysis and RNA degradation. The degradasome proteins are found in abundance in *B. subtilis* and since PfkA is a part of the degradasome, it has multiple functions (Cascante-Esteba et al., 2016). The glycolytic enzymes enolase and phosphofructokinase forms a complex known as glycosome, which may be involved in the increase of flux through glycolytic pathway (Newman et al., 2012). Deletion of *pfkA* can have affect on the levels of phosphorylated compounds in the cell and thus can influence the pyruvate pool (Muñoz-Márquez and Ponce-Rivas, 2010). Fructose-1, 6-bisphosphate which is a product of PfkA can contribute to the activity CcpA and CcpA acts as an activator for the *ilv-leu* operon by binding to its cre-site (Sonenshein, 2007) (Figure 52). So, in the absence of PfkA, there might be an alteration in the expression of *ilv-leu* operon due to the inactivation of CcpA and expression of this operon is essential for the biosynthesis of leucine, which is one of the most abundant precursor of surfactin within the peptidic moiety. Therefore, the deletion of *pfkA* has a dual repression affect on the surfactin production, firstly, lowering the flux towards pyruvate which is precursor for the

synthesis of branched chain amino acids and secondly, through inactivation of CcpA. Thus, it can be concluded that *pfkA* is essential for achieving both maximal growth rates and surfactin production.

The glycolytic enzymes Pfk and PykA play a crucial role in determining the carbon flux in the glycolytic pathway (Holger Ludwig et al., 2001). The gene *pykA* flanks to *pfkA* gene and both are expressed constitutively in *B. subtilis* (Muñoz-Márquez and Ponce-Rivas, 2010). This gene is responsible for the synthesis of pyruvate, which is the final step of glycolysis and pyruvate forms the principal metabolite linking several important pathways (Figure 44). Any of these pathways can be associated to cell cycle processes, therefore cells might use pyruvate as a fundamental metabolite in processes that regulates cell cycle and division with growth. Disruption of pyruvate kinase (*pyk*) can hinder the normal role of FtsZ protein and its association into Z ring, which can affect normal cell growth (Monahan et al., 2014). It was also hypothesized that in the absence of *pyk*, it can result in the accumulation of phosphoenol pyruvate (PEP), which can lead to lowering of growth rate (Fry et al., 2000). But in both cases, addition of pyruvate exogenously re-established the normal division in a *pyk* mutant strain. So it can be inferred that lowering of specific growth rate of BBG269 (BSB1, *sfp*⁺: *cm*^r, Δ *pyk*: *kan*^r) observed in classical Landy medium was due to one of the reasons or the accumulation of PEP interacts with FtsZ protein and repressing its activity. But this lowering of specific growth rate was not observed in AMBER medium due to the presence of malate in the nutrient composition and malate can be converted to pyruvate by the action of three enzymes, MaeA, MalS and MleA (Meyer and Stülke, 2013). It was found that the glucose consumption rate was delayed in *pyk* mutant. Transcript level analysis of several genes showed that there was an upregulation of genes related to non-PTS glucose transporters and down regulation of genes related biosynthesis of metabolites, which use pyruvate as precursor. There was an increase in transcript level of genes responsible for the conversion of malate to pyruvate to compensate the pyruvate limitation. It was also found that the transcript level of genes in the pentose phosphate pathway and gene for the conversion of glutamate to α -ketoglutarate was increased too (Cabrera-Valladares et al., 2012). In a *pyk* mutant due to accumulation of PEP, there was an increase in the production of various aromatic compounds i.e. shikimic acid, dehydroshikimic acid etc (Licon-Cassani et al., 2014). Thus, it can be concluded for these studies that the decline in surfactin production of BBG269 (BSB1, *sfp*⁺: *cm*^r, Δ *pyk*: *kan*^r) was majorly due to pyruvate limitation. But still gradual increase in

yield was observed in both classical Landy and AMBER medium at 6 hours and 8 hours was due to re-routing of flux through pentose-phosphate to pyruvate as well as synthesis of pyruvate through TCA cycle metabolites i.e. malate and α -ketoglutarate.

The above results proved that the prediction based on pyruvate influx is true and an increase in pyruvate flux can indirectly lead to an enhancement in surfactin production.

5. IMPACT OF DELETION ON SURFACTIN VARIANTS

Previously, all the studies were linked to surfactin's quantitative analysis but in this section the main focus is on the impact of deletion mutant on surfactin isoform or qualitative analysis of surfactin. Surfactin is a heptapeptide linked to a β -hydroxy group of fatty acid chain forming a cyclic lactone ring structure. The peptidic moiety is arranged in a cyclic sequence of Glu-Leu-Leu-Val-Asp-Leu-Leu with chirality LLDLLDL. The β -OH fatty acid length varies from 12 to 16 carbon atoms. Studies have revealed that the position 2, 4 and 7 of the peptidic moiety are generally substituted, out of which position 4 has the greatest substrate variation. The D-amino acids present in position 3 and 6 are highly conserved within the peptidic moiety (Table 3 page no.35). The lipidic tail of surfactin also exhibits heterogeneity. The β -OH fatty acid can be iso, anteiso C₁₃; iso, normal C₁₄ and iso, anteiso C₁₅ (Peypoux et al., 1999a). The chain lengths C₁₄ and C₁₅ are usually predominant (Jacques, 2011).

This structure is mainly responsible for various activities of this molecule. In addition to the main surfactin variant, the producing strain also produces different variant of surfactin but the concentration of this variant differs based on the genetic make-up of the strain. This variant production mostly depends on the amino acid present in the culture media or within the intracellular pool.

The *bkd* operon (*ptb-bcd-buk-lpdV-bkdAA-bkdAB-bkdB*) of *B. subtilis* encodes for enzymes responsible for the degradation of branched chain amino acids (valine, leucine and isoleucine) (Nickel et al., 2004). The promoter of the operon is present up-stream of *ptb* and the terminator is downstream of *bkdB*, thus encompassing the entire operon. Moreover, it was speculated that another promoter might be present within the coding region of *buk*. The gene *bkdR* is essential for the expression of *bkd* operon. The genes *lpdV*, *bkdAA*, *bkdAB* and *bkdB* encode E3, E1 α , E1 β and E2 proteins which are functional components of branched chain α -keto dehydrogenase (M Debarbouille et al., 1999).

5.1 Isoform analysis

Isoform analysis was carried out using LC-MS-MS with the 6 hour sample of each mutant (BBG258, BBG260, BBG261 and BBG262) as it has been described in materials and

methods (LC-MS-MS analysis of surfactin isomer, page no. 76). This time point was selected as at this time the *ilvB* operon (Nicolas et al., 2012) is highly expressed.

The variation in concentration of fatty acid chain length for two strains (BBG258 and BBG261) at 6 hours is shown in Figure 53. The percentage of C₁₄ was 32.2 and 67.2 for BBG258 and BBG261, respectively, whereas the percentage of C₁₅ was 56.9 and 24.9 for BBG258 and BBG261, respectively. The quantity of C₁₄ isoform was twice more in BBG261 in comparison to BBG258.

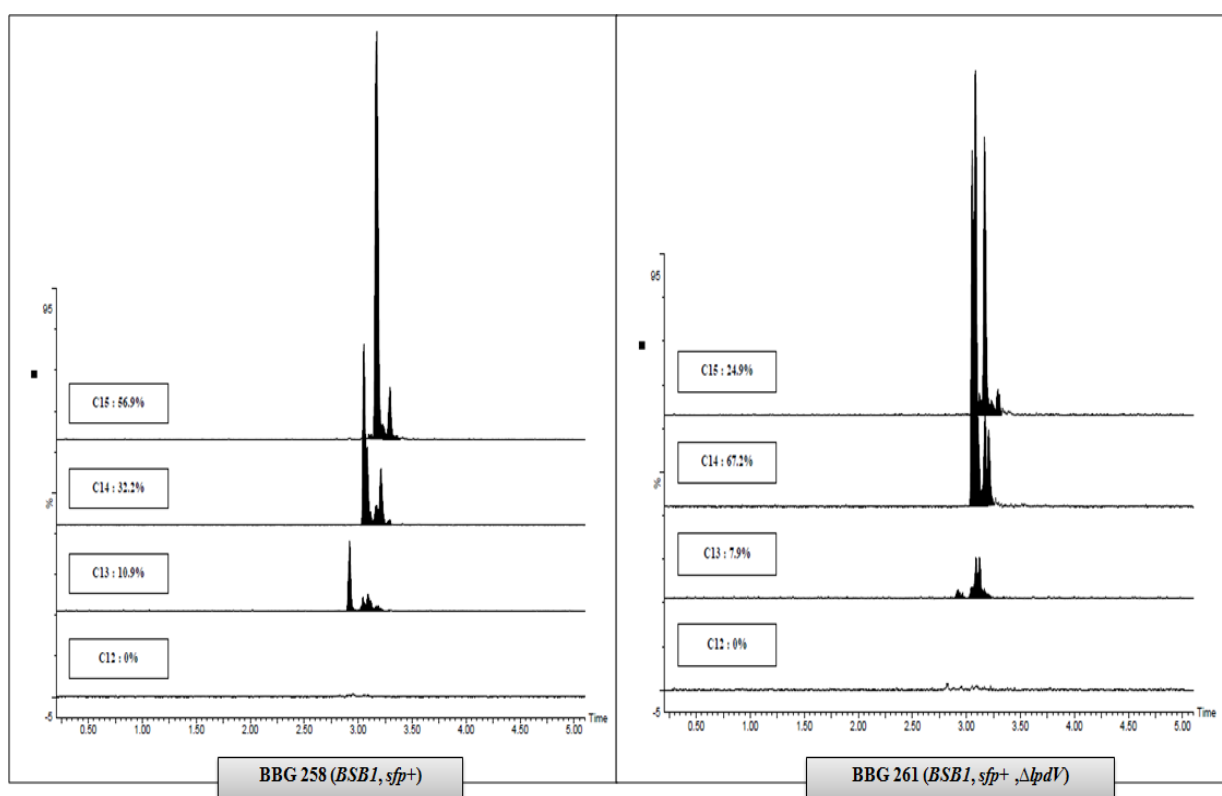


Figure 53 Concentration expressed as percentage of each fatty acid chain length in surfactin sample of BBG258 and BBG261

The two other mutants produced C₁₄ in less concentration in comparison to BBG261 (APPENDIX III). Thus a mutant strain was successfully developed which is a high producer of C₁₄ isoform. The role of overflow metabolites in the production of this isoform has been verified through extracellular metabolite analysis.

5.2 Extracellular metabolite analysis

Extracellular metabolites of the strains BBG258 and BBG261 were analyzed using nuclear magnetic resonance ($^1\text{H-NMR}$) spectroscopy at 310 K. Both standard compounds and Human Metabolome Database (HMDB) was applied for the identification of spectra while the quantification was carried out by exploiting the signal area of Quant Ref signal at 15 ppm (material and method: section 12). The strains were incubated in Landy medium with ammonium sulphate, buffered with phosphate, and the sampling was done at every 2 hour intervals upto 10 hours.

Samples were collected just after inoculation; this represents the concentration of the metabolites at 0 hour. The results of the extracellular depicts outflow of metabolite indicated by increase in concentration at a particular time into the extracellular medium whereas decrease in concentration at particular time indicates utilization or that metabolite was not produced.

As shown in Figure 54, the metabolic flux from pyruvate, flows to various pathways producing numerous metabolites, which are efflux into the extracellular medium. Pyruvate was also released into the medium and afterwards it was consumed for the production of various metabolites. The strain BBG258 efflux pyruvate with the concentration reaching upto 0.13 ± 0.009 mM at the end of 4 hours, whereas the maximum concentration of pyruvate in the medium by BBG261 was reached after 4 hours with a concentration 0.06 ± 0.007 mM. After 4 hours and 6 hours, the strains BBG258 and BBG261, respectively starts to consume pyruvate.

The release of high concentration of pyruvate by BBG258 into extracellular medium also indicates that this metabolite was produced in higher concentrations and may be directed into various pathways or released in higher concentration in comparison to BBG261.

After 6 hours, the strain BBG258 starts to produce lactate while the strain BBG261 did not directs the metabolic flux towards the synthesis of lactate indicated by almost constant concentration of lactate in the extracellular medium.

Both strains start to consume isoleucine in a similar pattern and after 5 hours and 6 hours, isoleucine is released into the medium by the strain BBG258 and BBG261 but in a concentration ranging from 0.01 to 0.03 mM.

Acetate was the main overflow metabolite produced by both strains. The maximum concentration produced by BBG258 and BBG261 was 17 ± 3.25 mM and 15 ± 3.5 mM after 8 hours, respectively.

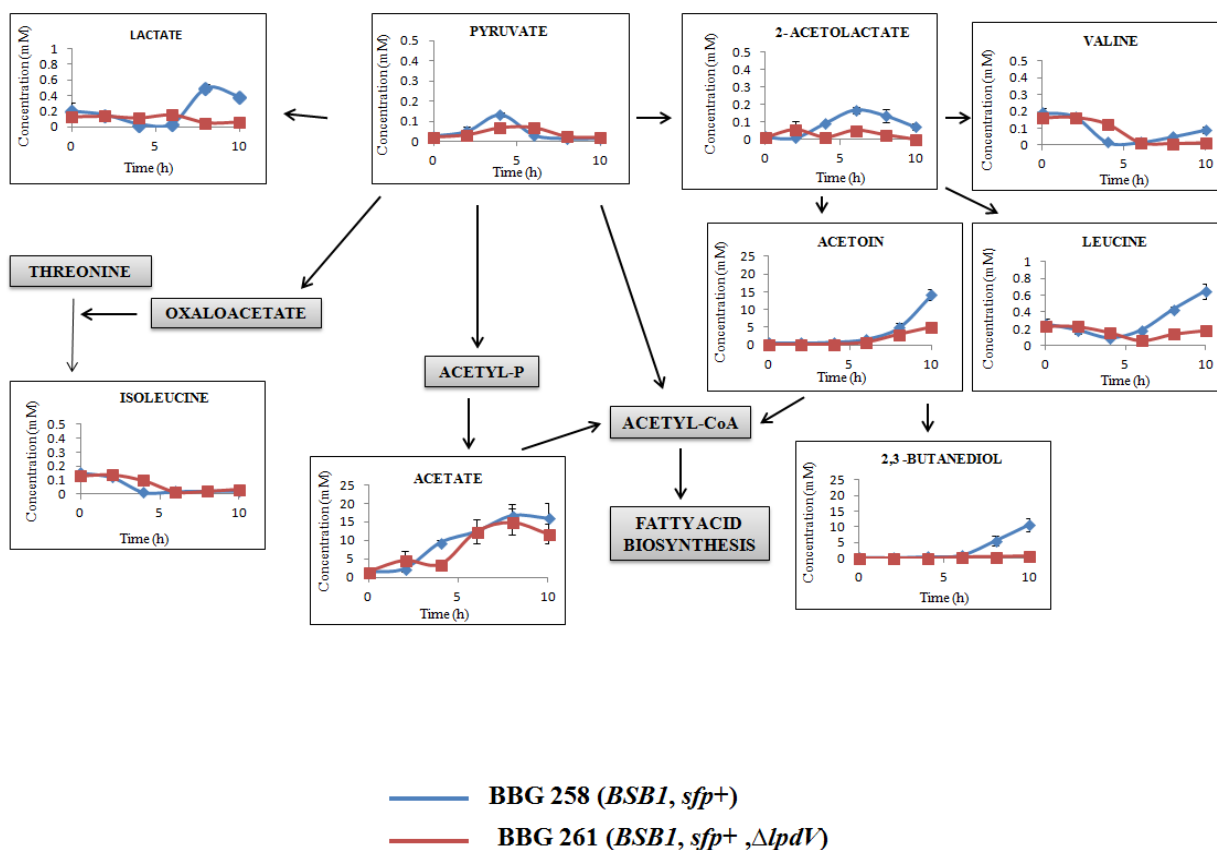


Figure 54 Time-resolved analysis of overflow metabolites and its associated metabolic pathways. Arrows indicate direct conversion, whereas metabolites in boxes indicate the intermediates. Results are expressed as mean value with standard deviation.

The metabolic flux from pyruvate is directed towards 2-acetolactate and from there, it is simultaneously converted to valine, leucine or acetoin. The concentration of acetolactate changes dynamically in the extracellular medium which indicates that it is produced and consumed for the synthesis of other metabolites. Similar pattern of production and consumption was followed by both strains and changes are in the order of 10^{-3} mM.

Acetoin was synthesized from 2-acetolactate. The maximum amount of flux was directed towards acetoin synthesis by BBG258 in comparison to BBG261. The maximum concentration in the medium reaching up to 14 ± 1.6 mM and 5 ± 0.7 mM for BBG258 and BBG261, respectively after 10 hours. It can be inferred that the metabolic flux in BBG261 was directed towards fatty acid synthesis and less towards acetoin synthesis in comparison to BBG258. Acetoin is converted to acetyl-coA, which is then used for the fatty acid biosynthesis. This hypothesis was also validated through the synthesis of 2, 3 – butandiol. The strain BBG258 produces 2, 3 – butandiol at a concentration of 11 ± 2.2 mM while a mere concentration of 0.7 ± 0.14 mM was produced by BBG261 at the end of 10 hours. The metabolic flux from 2-acetolactate is also directed towards the synthesis of valine and leucine.

Leucine is converted by a reversible reaction to 2-oxo-isocaproic acid by either of these genes *bcd*, *ywaA* and *ybgE*. Then the gene cluster *bkdAA* and *bkdAB* is responsible for the conversion of 2-oxo-isocaproic acid to 3-Methyl 1-hydroxybutyl and is simultaneously converted to 3-Methyl butanoyl-dihydrlipoamide-E by the same set of genes. This metabolite is converted to isovalerate and dihydrlipoamide-E by *bkdB*. The former is used for the synthesis of branched chain fatty acids (isoC15:0 or isoC17:0), while the latter is converted to lipoamide by *lpdV* and this is used as a substrate by the enzyme encoded by *bkdAA* and *bkdAB* for the synthesis of 3-Methyl butanoyl-dihydrlipoamide-E. The gene *lpdV* encoding for the enzyme E3 lipoamide dehydrogenase was deleted in the strain BBG261, the gene is shown in red (Figure 54).

As shown in Figure 54, leucine was released into the medium after 4 hours and 6 hours by BBG258 and BBG261, respectively. But the efflux concentration differs as maximum concentration in the medium reached by BBG258 was 0.6 ± 0.08 mM and for BBG261 it reached 0.18 ± 0.04 mM. The concentration of 2-oxo-isocaproic acid changed in both directions for BBG261 during the course of experiment, as it was released and consumed, while BBG258 releases this metabolite into the environment after 4 hours.

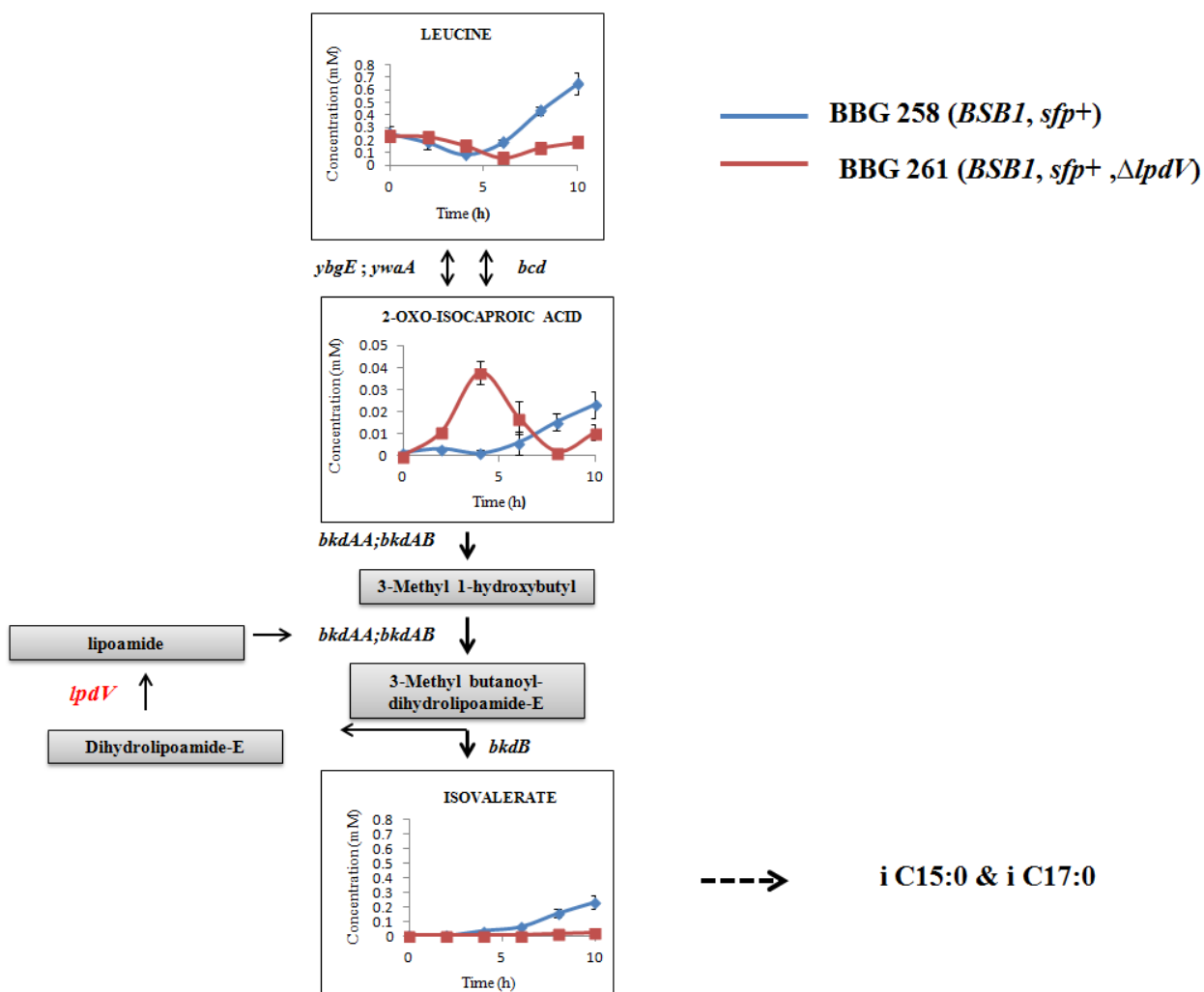


Figure 55 Extracellular concentration of leucine and its degradation products for BBG258 and BBG261, Dashed lines indicated conversion to respective fatty acid. Gene in red deleted. Results are expressed as mean value with standard deviation.

Isovalerate concentration in the medium reached up to a maximum value 0.23 ± 0.04 mM for BBG258, while for BBG261 the concentration in the medium remained constant until the end of cultivation.

Valine is converted by a reversible reaction to 3-methyl-2-oxobutyrate by either of these genes *bcd*, *ywaA* and *ybgE*. The gene cluster *bkdAA* and *bkdAB* is responsible for the conversion of 3-methyl-2-oxobutyrate to 2-Methyl 1-hydroxypropyl and is simultaneously converted to 2-Methyl propanoyl-dihydrlipoamide-E by the same set of genes. This metabolite is converted to isobutyrate and dihydrlipoamide-E by *bkdB*. The former is used for the synthesis of branched chain fatty acids (isoC14:0 or isoC16:0) while the latter is

converted to lipoamide by *lpdV* and this is used as a substrate by the enzyme encoded by *bkdAA* and *bkdAB* for the synthesis of 2-Methyl propanoyl-dihydrolipoamide-E.

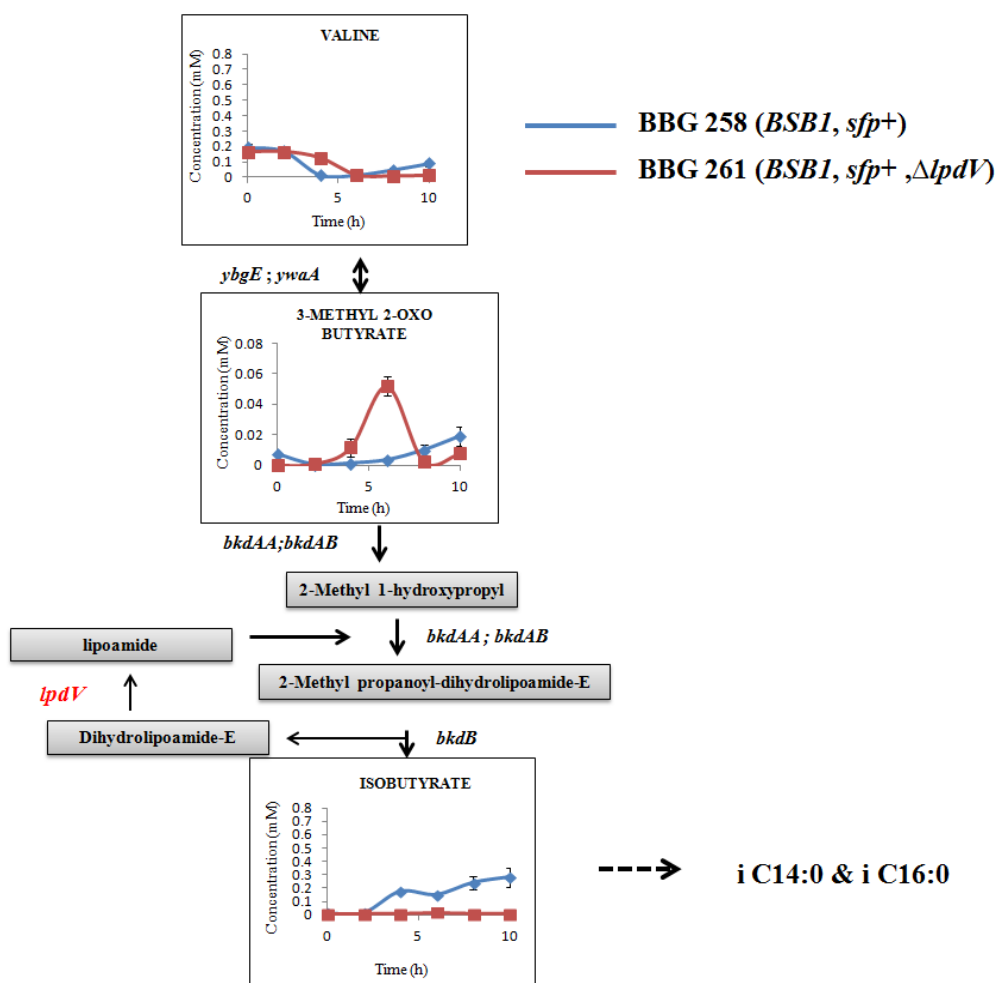


Figure 56 Extracellular concentration of valine and its degradation product for BBG258 and BBG261, Dashed lines indicates conversion to respective fatty acid. Gene in red deleted

The gene *lpdV* encoding for the enzyme E3 lipoamide dehydrogenase was deleted in the strain BBG261 (Figure 55). As shown in Figure 55, valine was released into the medium after 4 hours by BBG258, while it was completely consumed by BBG261. The efflux concentration in the medium reached by BBG258 was 0.08 ± 0.019 mM. The concentration of 3-methyl-2-oxobutyrate changed in both directions for BBG261 during the course of

environment as it was released and consumed while BBG258 releases this metabolite into the environment after 4 hours. Isobutyrate concentration in the medium reached up to a maximum value 0.28 ± 0.07 mM for BBG258 but for BBG261 the concentration in the medium remained constant until the end of cultivation.

Thus, it can be hypothesized from the results that the flux in BBG258 was mostly directed towards the production of overflow metabolites i.e. lactate, acetate, acetoin and 2, 3-butandiol and the branched chain acids (valine and leucine) was consumed in the initial stage and released into the environment in the later stage. But in the case of BBG261, the overflow metabolites were synthesized in less concentration except acetate in comparison to BBG258, as well as leucine and valine were not released into the medium. As a result, higher concentration of precursors are directed towards branched chain fatty acid synthesis from valine or leucine or towards conventional fatty acid synthesis from acetoin.

The strains were also analyzed for the metabolism of valine and leucine; it was observed that due to the deletion of *lpdV*, the recycling of dihydrolipoamide-E was inhibited leading to lower production of 2-Methyl propanoyl-dihydrolipoamide-E and 3-Methyl butanoyl-dihydrolipoamide-E with BBG261 in comparison to BBG258. This was well co-related by lower production or utilization of isobutyrate and isovalerate. This can also be speculated that these products are utilized by BBG261 for the production of branched chain fatty acids leading to production of isoforms of surfactin. The promoter of the *bkd* operon was reported to be induced by isoleucine and valine (M Debarbouille et al., 1999). The expression of this operon was reduced when the transporters responsible for the import of these two amino acids were deleted. It was observed that beside these two amino acids, leucine can also induce weaker activation of this promoter. The β -galactosidase activity of *ptb-lacZ* was observed 23 times and 17 times less with presence of leucine in comparison to isoleucine and valine respectively. Using *ptb-lacZ* fusion, it was observed that the promoter activity depends on the presence of amino acids in the following order isoleucine>valine>leucine (Belitsky et al., 2015).

Presence of valine serves as a precursor for the production of surfactin variants with even iso- β -hydroxy fatty acids whereas presence of isoleucine and leucine enhanced the production of surfactin variants with odd anteiso- and iso- β -hydroxy fatty acid, respectively. In the absence of any amino acids, the presence of β -hydroxy fatty acid in surfactin isoform was found to be $C_{15} > C_{14} > C_{13} > C_{16}$ (J. F. Liu et al., 2012). Similar results

were observed with BBG258 with no genetic alteration resulting in higher production of C₁₅ followed by C₁₄ and C₁₃, respectively.

It was found that both [Leu7] - and [Val7] - gave rise to C₁₄ surfactin variant although in the former variant, the iso C₁₄ was mixed with small amount anteiso form depicting that leucine can also give rise to a mixture of iso C₁₄ (Kowall et al., 1998). The concentration of leucine and isobutyrate present in extracellular medium at 6 hours was 0.18±0.02 mM and 0.15±0.009 mM for BBG258 whereas for BBG261 the concentration of the former and latter was 0.06±0.03 mM and 0.018±0.002 mM for leucine and isobutyrate at 6 hr, respectively (Figure 55 and Figure 56). Thus, it can be hypothesized from isoform analysis and surfactin analysis that in BBG261, deletion of *lpdV* directed the degradation of valine towards the synthesis of iso C₁₄. This was well supported by the concentration of leucine and isobutyrate in the extracellular medium and high percentage of C₁₄ through LC-MS analysis.

The production of this isoform in high concentrations can be used for industrial application, specifically as biosurfactant. Indeed, the spreading activity of the biosurfactant produced by *B. mojavensis* was enhanced when percentage of C₁₄ was more than C₁₅ as even-numbered fatty acid provide an optimum hydrophilic-lipophilic balance, which is an essential criterion for surface activity (Youssef et al., 2005).

5.3 Publication and Valorization of these results

ARTICLE

Dhali, D., Coutte, F., Arias, A.A., Auger, S., Bidnenko, V., Chataigné, G., Lalk, M., Niehren, J., Sousa, J., Versari, C., Jacques, P. (2017). **Genetic engineering of the branched fatty acid metabolic pathway of *Bacillus subtilis* for the overproduction of surfactin C₁₄ isoform.** *Biotechnology journal*, (submitted).

6. PERMEASE

Previously, in the section about transporters it was studied that *B. subtilis* contains three different types of transporters. Firstly, a set of transporter proteins function as importer of branched chain amino acids (valine, isoleucine and leucine). This set of proteins includes BcaP, BrnQ and BraB. The second set proteins contains that are involved in the export of branched chain amino acids, these include AzlC and AzlD (Figure 19). Lastly, it was proposed that a particular protein YvbW is a leucine specific permease (Wels et al., 2008). Moreover, it was recently hypothesized that YvbW might be a candidate for the transport of leucine specifically as it was observed that a quadruple mutant strain (*braB brnQ bcaP leuB*) was able to grow in Leu (40 µg/ml) containing minimal medium (Boris R Belitsky, 2015). As this gene was not included in the metabolic network (Figure 34), so this gene was not predicted among the single knockout prediction (Table 9). It can be hypothesized that this gene can have a role in increasing the intracellular pool of leucine. The surfactin production was studied both in the presence of native and constitutive promoter controlling *srfA* operon.

6.1 Mutant construction

The gene *yvbW* was deleted in the model organism *B.subtilis* BSB1 strain using Pop In – Pop Out technique (Figure 30). The genotype of the mutant strains are provided in Table 22.

Table 22 The mutants obtained with the deletion of *yvbW* along with the replacement of native promoter with constitutive promoter in *B. subtilis* strains

STRAIN	GENOTYPE	SOURCE
BBG258	<i>BSB1, sfp⁺: cm^r, neo^r</i>	This work
BBG257	<i>BS168, sfp⁺: cm^r, Δ<i>yvbW</i>, neo^r</i>	This work
BBG276	<i>BBG258, P_{repU}-erm^R::HaeIII-<i>srfA</i></i>	This work
BBG275	<i>BBG257, P_{repU}-erm^R::HaeIII-<i>srfA</i></i>	This work

The plasmid pBG401 harbours the P_{repU} promoter sequence cloned between the homologous region of the native P_{srfA} in *B. subtilis* 168 (Figure 28). The strains BBG258 and BBG257 were transformed with this plasmid through natural transformation technique and selected on LB-erythromycin. The integration of P_{repU} promoter was verified through PCR. The genotype of the obtained mutants has been shown in Table 22.

6.2 Growth and surfactin analysis with native promoter and constitutive promoter

The mutants obtained with *B. subtilis* BSB1 (BBG257) along with control strain (BBG258) was cultivated in Landy medium with ammonium sulphate and phosphate buffer condition (APPENDIX II).

The experiment was performed in a shake flask at 37°C and 250 rpm. Suicidal flask strategy was used i.e. one flask for each time point (6 h, 10 h and 24 h). Experiments were carried out in triplicate and two biological replicates were performed. Results are expressed as mean value with standard deviation.

Similar strategy was used for the analysis of growth and surfactin production for the strains BBG275 and BBG276. But with these strains, the samples were collected only at 6 h and 10 h.

6.2.1 Strains with native promoter (P_{srfA})

The growth of the strains (BBG257 and BBG258) was analysed by measuring the optical density at 600 nm at each time point. Overview of the specific growth rate of all the strains

is provided in Table 22. Growth kinetics study of all the strains was carried out during the exponential growth phase.

The specific growth rate of BBG258 in Landy medium with MOPS as buffering agent was found to be 0.42 ± 0.01 (Table 16) but the specific growth of BBG258 was lowered in this condition and was found to be 0.37 ± 0.01 . The reason for such a behaviour is due to lower concentration of magnesium ions in this media in comparison to Landy medium with MOPS. Moreover this decline in concentration did not affect the growth rate of BBG257 in comparison to BBG258 in Landy medium with MOPS.

Table 23 Specific growth rate of BBG258 and BBG257 in modified Landy medium buffered with phosphate buffer

STRAINS	SPECIFIC GROWTH RATE (h^{-1})
BBG258 (BSB1, <i>sfp</i> ⁺ : cm ^r , neo ^r)	0.37 ± 0.01
BBG257 (BS168, <i>sfp</i> ⁺ : cm ^r , $\Delta yvbW$)	0.501 ± 0.018

Since most of the genes that regulates the branched chain amino acid metabolic pathway in *B. subtilis* are pleiotropic in nature, the effect of deletion on the growth of the strains was studied. Growth analysis revealed that the deletion did not affect the growth rate of the mutant strain.

Surfactin production was analyzed with the strains in this media condition at different time intervals. Both strains produced surfactin but with variable concentrations (Figure 57). The surfactin production for BBG258 was found to be 68 ± 5 mg/l, 220 ± 83 mg/l and 694 ± 93 mg/l at 6 h, 10 h and 24 h, respectively, while the surfactin production for BBG257 was found to be 632 ± 89 mg/l, 2453 ± 254 mg/l and 4838 ± 114 mg/l at 6 h, 10 h and 24 h, respectively. The maximum surfactin production was found at 24 hour for both strains. At this time point, BBG257 produced 7 times more than BBG258.

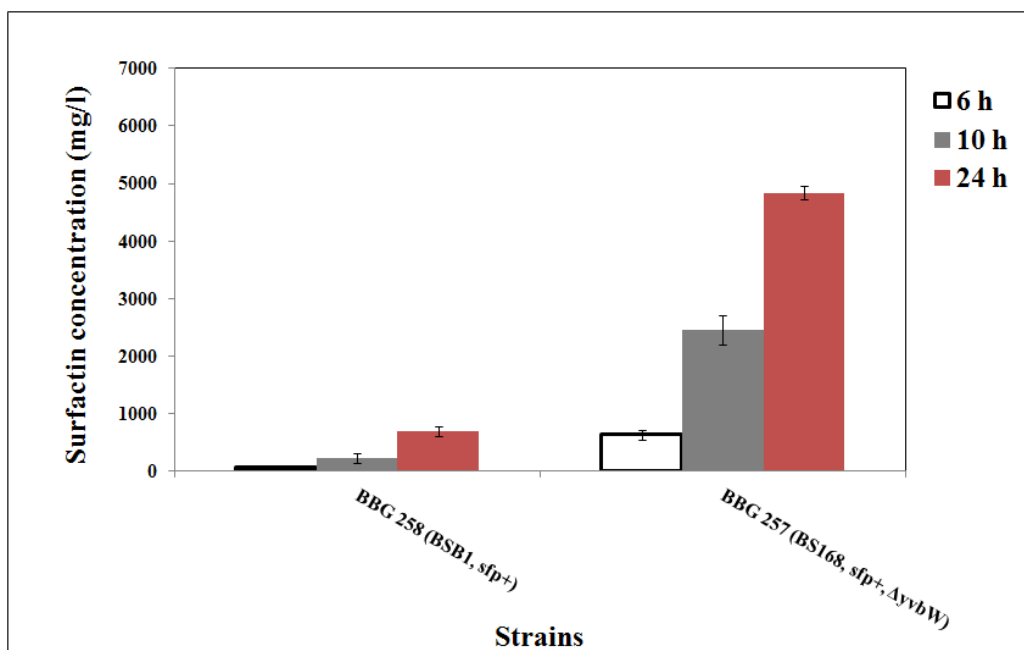


Figure 57 Surfactin production (mg/l) by BBG258 and BBG257 in Landy medium with phosphate buffer at 37°C with 250 rpm in shake flask condition. Results are expressed as mean value with standard deviation.

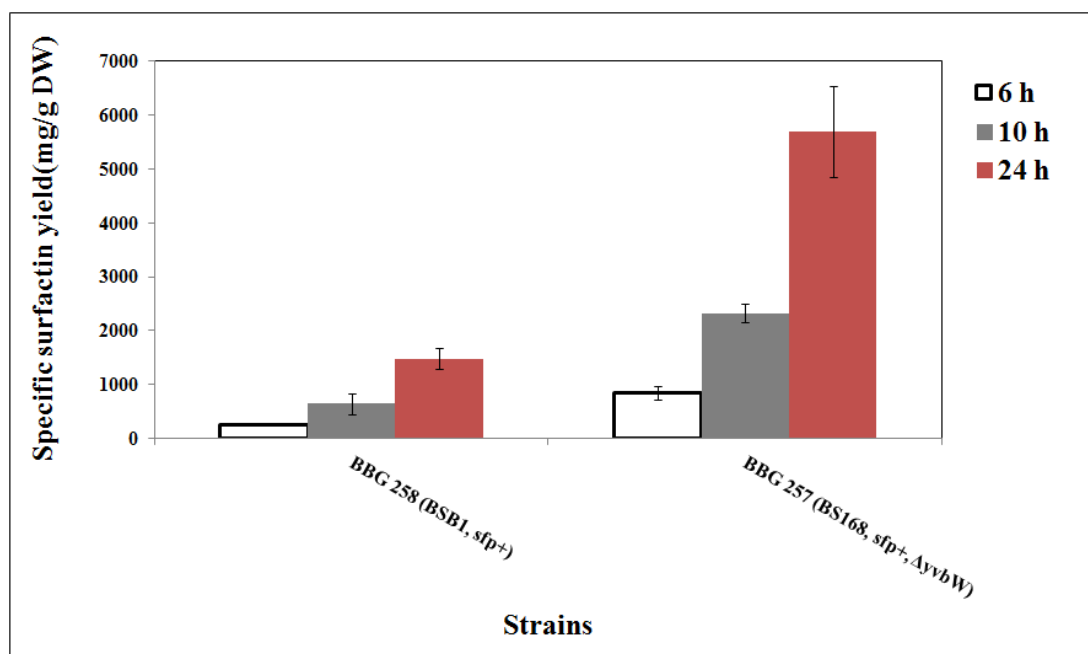


Figure 58 Specific surfactin yield obtained with BBG258 and BBG257 in Landy medium with phosphate buffer. Results are expressed as mean value with standard deviation.

The surfactin concentration was normalized with the cell biomass dry weight (DW) of each strain at each time point to obtain the specific surfactin yield which is represented by mg of surfactin per gm of dry weight. The specific surfactin yield obtained with BBG258 was 257 ± 11 ; 639 ± 203 ; 1480 ± 195 mg of surfactin per gm of DW whereas with BBG257 it was obtained 839 ± 126 ; 2324 ± 174 ; 5696 ± 842 mg of surfactin per gm of DW at 6 h , 10 h and 24 h, respectively (Figure 58). At 24 hours, the specific surfactin yield was found to be 4 folds higher with BBG257 in comparison to BBG258.

6.2.2 Strains with constitutive promoter (P_{repU})

The growth of the strains (BBG275 and BBG276) was analysed by measuring the optical density at 600 nm at each time point. Overview of the specific growth rate of all the strains is provided in Table 24. Growth kinetics study of all the strains was carried out during the exponential growth phase.

Table 24 Specific growth rate of BBG258 and BBG257 in modified Landy medium buffered with phosphate buffer

STRAINS	SPECIFIC GROWTH RATE (h^{-1})
BBG276 (BBG258, P_{repU} -erm ^R :: P_{srfA})	0.397 ± 0.002
BBG275 (BBG257, P_{repU} -erm ^R :: P_{srfA})	0.501 ± 0.007

The specific growth rate of BBG258 and BBG257 in this medium was found to be 0.37 ± 0.01 and 0.501 ± 0.018 (Table 24) and the specific growth rate of BBG276 and BBG275 was found to be 0.397 ± 0.002 and 0.501 ± 0.007 , respectively. The specific growth rate of the strains did not change significantly, when the native surfactin promoter (P_{srfA}) was exchanged with a constitutive one (P_{repU}), controlling the expression of surfactin operon ($srfA$).

Surfactin production was analyzed with the strains in this media condition at 6 hours and 10 hours. Both strains produced surfactin but with diverse concentrations (Figure 59). The surfactin production for BBG276 was found to be 134 ± 18 mg/l and 293 ± 103 mg/l at 6 h

and 10 h, respectively. While the surfactin production for BBG275 was found to be 1048 ± 47 mg/l and 1511 ± 93 mg/l at 6 h and 10 h, respectively. The maximum surfactin production was found at 10 hours for both strains. At this time point, BBG276 produced 5 times surfactin more than BBG275.

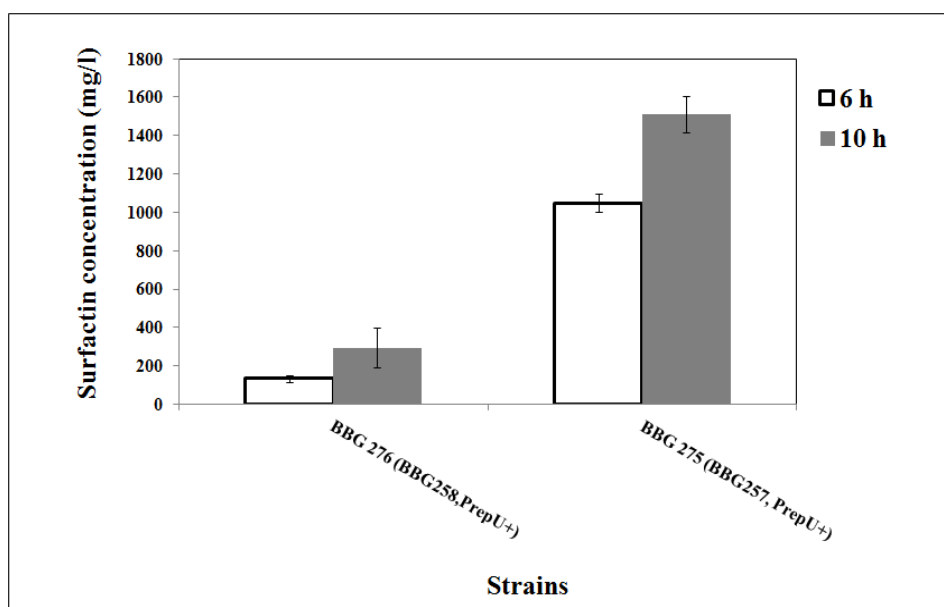


Figure 59 Surfactin production (mg/l) by BBG276 and BBG275 in Landy medium with phosphate buffer at 37°C with 250 rpm in shake flask condition. Results are expressed as mean value with standard deviation.

The surfactin concentration was normalized with the cell biomass dry weight of (DW) of each strain at each time point to obtain the specific surfactin yield in mg of surfactin per g of dry weight (Figure 60). The specific surfactin yield obtained with BBG276 was 369 ± 101 ; 521 ± 142 mg of surfactin per g of DW whereas with BBG275 it was obtained 1076 ± 142 ; 1143 ± 66 mg of surfactin per g of DW at 6 h and 10 h respectively. At 10 hours the specific surfactin yield of BBG275 was twice higher than that of BBG276.

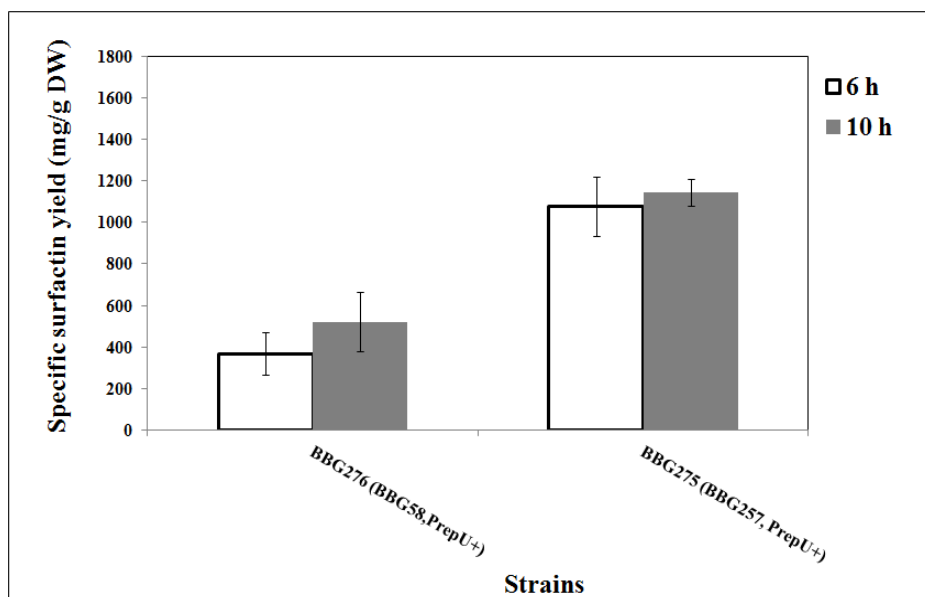


Figure 60 Specific surfactin yield obtained with BBG258 and BBG257 in Landy medium with phosphate buffer. Results are expressed as mean value with standard deviation.

Two major conclusions can be drawn from the above results; Firstly, the surfactin native promoter (P_{srfA}) of *B. subtilis* is a highly efficient promoter for the production of surfactin in comparison to the constitutive P_{repU} promoter. As BBG257 harboring the surfactin native promoter enhanced surfactin yield twice in comparison to its counterpart BBG275, having constitutive promoter at 10 hours. On the basis of native promoter controlling *srfA* expression, it was proved that P_{srfA} of *B. subtilis* 168 was stronger than P_{srfA} of *B. subtilis* ATCC6633 (Duitman et al., 2007). But in comparison to an other study, utilizing P_{repU} promoter for surfactin production (F. Coutte et al., 2010), the specific surfactin yield with BBG113 (*BS168*, sfp^+ , P_{repU} -*srfA*) was 880 mg of surfactin per g of DW at 24 hour, which is 1.3 times less than the yield obtained with BBG275 at 10 hour although the former study was carried out in Landy MOPS medium with glutamic acid and 130 rpm, while the latter was carried out in Landy with ammonium sulphate with phosphate buffer at 250 rpm. Although the constitutive promoter (P_{repU}) gave 3-folds and 15-folds higher production of iturin A (K Tsuge et al., 2001) and mycosubtilin (Leclère et al., 2005) in comparison to corresponding strain with native promoter. Another constitutive promoter P_{veg} was inserted into *B. subtilis* 3A38 to enhance the production of surfactin but a maximum production of 260 mg/l of surfactin was achieved as well as there was a decline in surfactin production

when this promoter was inserted into *B. subtilis* DSM 10^T strain. The structural organization of native promoter revealed the presence of dyad symmetries upstream of ribosomal binding site of *urfA* operon, these symmetries are site for ComA binding (Willenbacher et al., 2016). This might be the reason for the stronger expression of P_{urfA} in relation to constitutive promoters (P_{repU} and P_{veg}) as ComA positively regulates *urfA* expression through its phosphorylated product (Figure 23).

Secondly, the gene *yvbW* is an interesting candidate for the overproduction of surfactin. The strain BBG257 (*BSB1*, *sfp*⁺, Δ *yvbW*) gave the highest yield of surfactin of 5.7 g per g of cell biomass, among the mutant strains screened in this study as well as all the strains reported in literature, in shake flask condition and without foam fractionation. It was found that this strain was able to convert 37% carbon of 20 g/l glucose to the carbon required for surfactin biosynthesis in shake flask condition. So, it can be hypothesized that upon deletion of this gene the export of leucine was inhibited upto certain extent and more leucine molecules are available as precursors for the surfactin synthetase for production of surfactin.

The T-box anti-termination mechanism is highly sensitive and allows the bacteria to maintain steady level of aminoacyl tRNA (Merino et al., 2005). The T-box controls genes that are involved in amino acid biosynthesis and transport. Most of the genes involved in the transport of amino acid are regulated by T-box mechanism. The gene *yvbW* found in *B. subtilis* has leucine specific T-box located upstream of the gene. This gene belongs to APC family of transporters (Wels et al., 2008). Under both high and limiting leucine condition, there is expression of *yvbW* but the level is higher in former condition and lower in later condition (Rollins, 2014).

6.3 Publication and Valorization of these results

PATENT

Dhali, D., Coutte, F., Jacques P. Biosurfactant of *B. subtilis* : overproducing strains and genetic engineering methods to obtain (*to be submitted*)

CHAPTER 4: CONCLUSION AND PERSPECTIVES

CONCLUSION

The study started with a hypothesis that an increase in surfactin production can be obtained with an enhancement of precursor concentration. The reason behind this speculation was that previous studies revealed that the monoproducer lipopeptide strain as well as replacing the promoter did not enhance the surfactin yield. The peptidic moiety of surfactin contains four residues of leucine, so the strain was cultivated in the presence/absence of leucine (Figure 61-1). It was observed that the strain upon leucine feeding produced maximum 1.8 g of surfactin per g of cell biomass. This yield was 3 times higher than the control condition where there was no leucine. This result proved that leucine is one of the precursors required for increasing the production of surfactin or has an indirect role through the regulation of its biosynthesis. The next step was to have the intensive knowledge regarding the metabolism of leucine in *B. subtilis*. But the entire knowledge regarding the regulation and metabolism is not fully studied until now. Therefore, an integrated approach was required. A complete model of one of the most complex and regulated metabolic pathway of *B. subtilis*, i.e. the branched chain amino acid metabolic pathway was developed in collaboration with the Biocomputing team of CRISAL laboratory Lille. This network constitutes the regulation and metabolism of the branched chain amino acids (isoleucine, valine and leucine) (Figure 61-2). This model was established using computational language. Based on this metabolic network, 12 plausible single knockout predictions were returned as well as 2 predictions regarding flux of metabolites in order to improve leucine intracellular pool. Out of 12 single knockout predictions, 6 single knockouts was carried out in *B. subtilis* 168 strain and 4 similar knockouts were carried out again in *B. subtilis* BSB1 strain. In Landy medium, with ammonium sulphate (Figure 61-2), the maximum specific surfactin yield was obtained with *codY* deletion was 1.3 g and 1.27 with the strains BBG254 and BBG260, respectively. The deletion of *codY* has the maximum significant impact on surfactin yield as the yield was enhanced 7-fold and 5-fold in comparison to their respective control strains. This was coherent with the metabolic network information as from the metabolic network it was observed that CodY can repress leucine synthesis through various repression mechanism. The specific yield of *TnrA* mutants (BBG256 and BBG262) derived from *B. subtilis* 168 and *B. subtilis* BSB1 was found to be significant in comparison to control but the yield was lower in comparison to respective *codY* mutant strains. But the yield obtained with the

double mutant of *codY* and *TnrA* (BBG265) was lower in comparison to *codY* and *TnrA* single knock-out strains. The reason for such observation was due to the fact that *TnrA* acts as an activator for the transport of ammonium. There was a significant impact on enhancement of surfactin production due to deletion of *bcd* gene as this gene is responsible for the degradation of leucine into its specific fatty acid. These set (Figure 61-2) of results validated the prediction provided from the metabolic network and that there is a definitive role of leucine in enhancing surfactin production.

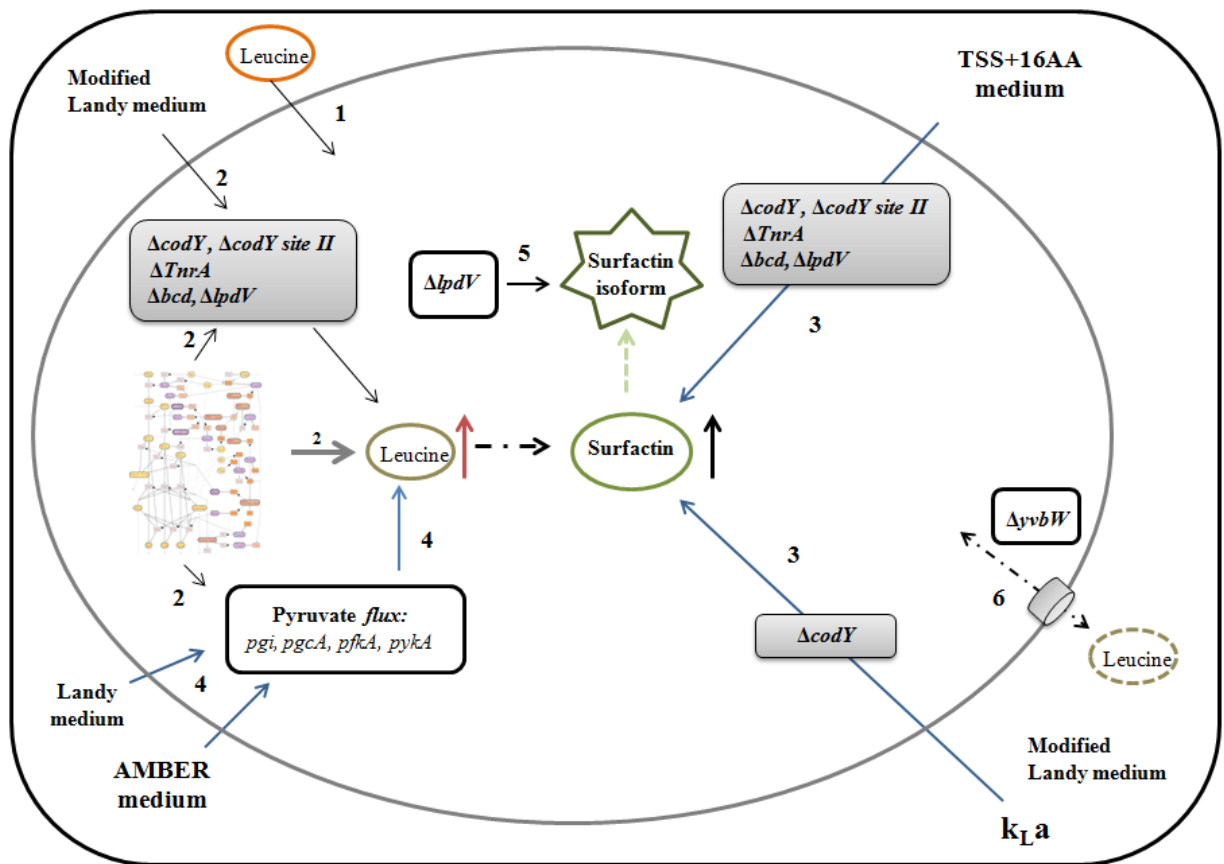


Figure 61 Overview of approaches to determine the role of precursors for the enhancement of surfactin in *B. subtilis*

The dependence of surfactin yield on various other factors i.e. glucose concentration, volumetric oxygen transfer rate and genetic construction was analyzed (Figure 61-3). It was found that in Landy medium with high glucose concentration the surfactin yield obtained was higher in comparison to TSS medium supplemented with 16 amino acids. Besides low glucose concentration, the presence of amino acids (valine and isoleucine)

activates the CodY repression on *ilv-leu* and *srfA* operon. This repression was reflected on low surfactin yield in all mutants, except for the mutants with *codY* and *codY* site II deletion. The volumetric oxygen transfer coefficient is a key parameter for surfactin production and the selectivity of the parameters determining the aeration (Figure 61-3). Therefore, to scale-up the fermentation process from flasks to large-scale fermentors, this parameter should be considered. Surfactin yield is clearly favoured at good oxygenation of the cells ($199 \text{ h}^{-1} > k_{\text{L}}a < 400 \text{ h}^{-1}$). Thus, there is an upper limit until which $k_{\text{L}}a$ has a linear relationship with specific surfactin yield, after that the relationship depends on the genetic construction of the strain. This parameter also depends on the genotype of the strain, as in this study it was shown that the maximum surfactin yield of BBG258 did not decrease at high $k_{\text{L}}a$ value but the yield of BBG260 was affected when this value was enhanced beyond 400 h^{-1} . The promoter replacement did not have great impact on the surfactin yield (Figure 61-3). The surfactin yield obtained with the strains harboring native promoter was higher in comparison to the strains having constitutive promoter at 6 hours. But at 10 hours, the surfactin yield from strains with either promoter was found to be equal.

Impact on pyruvate flux was analyzed with deletion of genes in the glycolytic pathway (Figure 61-4). All the mutants except $\Delta pgcA$ showed significant decrease in specific surfactin yield in comparison to control strain (BBG258). The maximum deterioration in surfactin yield was obtained with *pfkA* deletion mutant as the product of this gene can increase pyruvate flux but also can activate CcpA protein, which is essential for *ilv-leu* operon to inhibit binding of CodY. It was also observed that in the presence of glutamic acid in Landy medium, the specific surfactin yield of BBG258 was enhanced in comparison to Landy medium with ammonium sulphate.

Besides quantitative analysis of surfactin, qualitative analysis was also carried out (Figure 61-5). It showed that the *lpdV* mutant resulted in greater production of C₁₄-isoform in comparison to other mutant. This could be due to the fact that the flux was more directed towards the production of branched chain fatty acids or normal fatty acid synthesis. This hypothesis was verified through the analysis of extracellular metabolites. The decline in concentration of acetoin, isovalerate and isobutyrate in comparison to control strain supports the fact that the flux was directed towards fatty acid synthesis.

A positive influence of the knockout of a protein with a hypothetical function as leucine permease was verified (Figure 61-6). A specific yield of 5.7 g per g of DW was observed, which is reported to be the highest in shake flask condition without foam fractionation.

The deletion of this mutant directed 37% of the carbon in glucose towards carbon synthesis of surfactin molecule. But when the promoter was replaced with a constitutive one, there was only 2-fold enhancement in comparison to its counterpart and yield obtained was 1.14 g per g DW. The yield obtained with native promoter with the deletion of *yvbW* was 5-times more in comparison to strain with constitutive promoter with same gene deletion. This difference in yield might be due to the presence of dyad symmetries in the upstream region of native promoter for ComA binding.

Thus, the precursor limitation can be overcome with this integrative approach and can lead to high production of surfactin, which is an essential criteria for its proper industrial application.

PERSPECTIVES

1. Development of a double mutant ($\Delta codY$, $\Delta yvbW$) and analyze the yield of this mutant.
2. To analyze extracellular and intracellular metabolites for the overproducing mutant strains in the definitive conditions.
3. Extension of the modeling to all other metabolism, in order to have an integrative approach on the entire metabolism.
4. To study whether there might be a role of any other amino acid or fatty acid as a limiting factor for the production of surfactin.

REFERENCES

- Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z.N., Barabási, A.-L., 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* 427, 839–43. doi:10.1038/nature02289
- Amin, G.A., 2014. Exponential fed-batch strategy for enhancing biosurfactant production by *Bacillus subtilis*. *Water Sci. Technol.* 70, 234–40. doi:10.2166/wst.2014.211
- Antelmann, H., Darmon, E., Noone, D., Veening, J.-W., Westers, H., Bron, S., Kuipers, O.P., Devine, K.M., Hecker, M., van Dijl, J.M., 2003. The extracellular proteome of *Bacillus subtilis* under secretion stress conditions. *Mol. Microbiol.* 49, 143–56.
- Arima, K., Kakinuma, A., Tamura, G., 1968. Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation. *Biochem. Biophys. Res. Commun.* 31, 488–94.
- Bais, H.P., Fall, R., Vivanco, J.M., 2004. Biocontrol of *Bacillus subtilis* against infection of *Arabidopsis* roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiol.* 134, 307–19. doi:10.1104/pp.103.028712
- Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I., Mé digue, C., Danchin, A., Antoine Danchin antoinedanchin, C., 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. doi:10.1099/mic.0.027839-0
- Becher, D., Büttner, K., Moche, M., Hessling, B., Hecker, M., 2011. From the genome sequence to the protein inventory of *Bacillus subtilis*. *Proteomics* 11, 2971–80. doi:10.1002/pmic.201100090
- Belitsky, B.R., 2015. Role of branched-chain amino acid transport in *Bacillus subtilis* CodY activity. *J. Bacteriol.* 197, 1330–8. doi:10.1128/JB.02563-14
- Belitsky, B.R., 2015. Role of branched-chain amino acid transport in *Bacillus subtilis* CodY activity. *J. Bacteriol.* 197, 1330–1338. doi:10.1128/JB.02563-14
- Belitsky, B.R., Brinsmade, S.R., Sonenshein, A.L., 2015. Intermediate Levels of *Bacillus subtilis* CodY Activity Are Required for Derepression of the Branched-Chain Amino Acid Permease, BraB. *PLoS Genet.* 11, 1–22. doi:10.1371/journal.pgen.1005600
- Belitsky, B.R., Gustafsson, M.C., Sonenshein, A.L., Von Wachenfeldt, C., 1997a. An lrp-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. *J. Bacteriol.* 179, 5448–57.
- Belitsky, B.R., Gustafsson, M.C., Sonenshein, a L., Von Wachenfeldt, C., 1997b. An lrp-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. *J. Bacteriol.* 179, 5448–5457.

- Belitsky, B.R., Sonenshein, A.L., 2011. CodY-mediated regulation of guanosine uptake in *Bacillus subtilis*. *J. Bacteriol.* 193, 6276–87. doi:10.1128/JB.05899-11
- Belitsky, B.R., Sonenshein, A.L., 2011. Contributions of multiple binding sites and effector-independent binding to cody-mediated regulation in *bacillus subtilis*. *J. Bacteriol.* 193, 473–484. doi:10.1128/JB.01151-10
- Belshaw, P.J., Walsh, C.T., Stachelhaus, T., 1999. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* 284, 486–9.
- Biochem, J., Suzuki, K., Imahori, K., 1974. Phosphoglycerate Kinase of *Bacillus stearothermophilus* 2. The enzyme was composed of a single polypeptide chain of molecular weight 42,000. The thermophilic enzyme had many properties in common with enzymes from mesophilic sources, e.g., pH optimum 7.6, 771–782.
- Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–62.
- Bonmatin, J.-M., Labb, H., Grangemard, I., Peypoux, F., Maget-Dana, R., Ptak, M., Michel, G., 1995. Production, isolation and characterization of [Leu4]- and [Ile4]surfactins from *Bacillus subtilis*. *Lett. Pept. Sci.* 2, 41–47. doi:10.1007/BF00122922
- Bonmatin, J.-M., Laprévotte, O., Peypoux, F., 2003. Diversity among microbial cyclic lipopeptides: iturins and surfactins. Activity-structure relationships to design new bioactive agents. *Comb. Chem. High Throughput Screen.* 6, 541–56.
- Bonmatin, J.M., Genest, M., Labbé, H., Ptak, M., 1994. Solution three-dimensional structure of surfactin: a cyclic lipopeptide studied by ¹H-NMR, distance geometry, and molecular dynamics. *Biopolymers* 34, 975–86. doi:10.1002/bip.360340716
- Braaz, R., Wong, S.-L., Jendrossek, D., 2002. Production of PHA depolymerase A (PhaZ5) from *Paucimonas lemoignei* in *Bacillus subtilis*. *FEMS Microbiol. Lett.* 209, 237–41.
- Brans, A., Filée, P., Chevigné, A., Claessens, A., Joris, B., 2004. New integrative method to generate *Bacillus subtilis* recombinant strains free of selection markers. *Appl. Environ. Microbiol.* 70, 7241–50. doi:10.1128/AEM.70.12.7241-7250.2004
- Brinsmade, S.R., Alexander, E.L., Livny, J., Stettner, A.I., Segrè, D., Rhee, K.Y., Sonenshein, A.L., 2014. Hierarchical expression of genes controlled by the *Bacillus subtilis* global regulatory protein CodY. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2–7. doi:10.1073/pnas.1321308111
- Brinsmade, S.R., Kleijn, R.J., Sauer, U., Sonenshein, A.L., 2010. Regulation of CodY activity through modulation of intracellular branched-chain amino acid pools. *J. Bacteriol.* 192, 6357–6368. doi:10.1128/JB.00937-10

- Brinsmade, S.R., Sonenshein, A.L., 2011. Dissecting complex metabolic integration provides direct genetic evidence for CodY activation by guanine nucleotides. *J. Bacteriol.* 193, 5637–48. doi:10.1128/JB.05510-11
- Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., Stubbs, M.T., 2002. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* 10, 301–310. doi:10.1016/S0969-2126(02)00716-5
- Bu, J., Funke, M., Diederichs, S., Kensy, F., Mu, C., 2009. The Baffled Microtiter Plate : Increased Oxygen Transfer and Improved Online Monitoring in Small Scale Fermentations 103, 1118–1128. doi:10.1002/bit.22341
- Buescher, J.M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., Hessling, B., Kleijn, R.J., Le Chat, L., Lecointe, F., Mäder, U., Nicolas, P., Piersma, S., Rügheimer, F., Becher, D., Bessieres, P., Bidnenko, Sauer, U., 2012. Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science* 335, 1099–103. doi:10.1126/science.1206871
- Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P., Kucherov, G., 2008. NORINE: a database of nonribosomal peptides. *Nucleic Acids Res.* 36, D326-31. doi:10.1093/nar/gkm792
- Cabrera-Valladares, N., Martinez, L.M., Flores, N., Hernandez-Chvez, G., Martinez, A., Bolvar, F., Gosset, G., 2012. Physiologic consequences of glucose transport and phosphoenolpyruvate node modifications in *Bacillus subtilis* 168. *J. Mol. Microbiol. Biotechnol.* 22, 177–197. doi:10.1159/000339973
- Carrillo, C., Teruel, J.A., Aranda, F.J., Ortiz, A., 2003. Molecular mechanism of membrane permeabilization by the peptide antibiotic surfactin. *Biochim. Biophys. Acta* 1611, 91–7.
- Cascante-Esteva, N., Gunka, K., Stülke, J., 2016. Localization of Components of the RNA-Degrading Machine in *Bacillus subtilis*. *Front. Microbiol.* 7, 1–9. doi:10.3389/fmicb.2016.01492
- Cheng, J., Guan, C., Cui, W., Zhou, L., Liu, Z., Li, W., Zhou, Z., 2016. Corresponding author : Zhemin Zhou and Weijiang Li. *Protein Expr. Purif.* doi:10.1016/j.pep.2016.07.008
- Chollet-Imbert, M., Gancel, F., Slomianny, C., Jacques, P., 2009. Differentiated pellicle organization and lipopeptide production in standing culture of *Bacillus subtilis* strains. *Arch. Microbiol.* 191, 63–71. doi:10.1007/s00203-008-0429-8
- Commichau, F.M., Pietack, N., Stülke, J., 2013. Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Mol. Biosyst.* 9, 1068–75. doi:10.1039/c3mb25595f
- Commichau, F.M., Stülke, J., 2008. Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. *Mol. Microbiol.* 67, 692–702. doi:10.1111/j.1365-2958.2007.06071.x

- Commichau, F.M., Stülke, J., 2008. Trigger enzymes: Bifunctional proteins active in metabolism and in controlling gene expression. *Mol. Microbiol.* 67, 692–702. doi:10.1111/j.1365-2958.2007.06071.x
- Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G., van Sinderen, D., 1993. Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol. Microbiol.* 8, 821–831. doi:10.1111/j.1365-2958.1993.tb01629.x
- Coutte, F., Leclère, V., Béchet, M., Guez, J.S., Lecouturier, D., Chollet-Imbert, M., Dhulster, P., Jacques, P., 2010. Effect of pps disruption and constitutive expression of *srfA* on surfactin productivity, spreading and antagonistic properties of *Bacillus subtilis* 168 derivatives. *J. Appl. Microbiol.* 109, 480–491. doi:10.1111/j.1365-2672.2010.04683.x
- Coutte, F., Lecouturier, D., Yahia, S.A., Leclere, V., Bechet, M., Jacques, P., Dhulster, P., 2010. Production of surfactin and fengycin by *Bacillus subtilis* in a bubbleless membrane bioreactor. *Appl. Microbiol. Biotechnol.* 87, 499–507. doi:10.1007/s00253-010-2504-8
- Coutte, F., Niehren, J., Dhali, D., John, M., Versari, C., Jacques, P., 2015. Modeling leucine's metabolic pathway and knockout prediction improving the production of surfactin, a biosurfactant from *Bacillus subtilis*. *Biotechnol. J.* 10, 1216–1234. doi:10.1002/biot.201400541
- Crosa, J.H., Walsh, C.T., 2002. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev.* 66, 223–49. doi:10.1128/MMBR.66.2.223-249.2002
- Dalbey, R.E., Wang, P., van Dijl, J.M., 2012. Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol. Mol. Biol. Rev.* 76, 311–30. doi:10.1128/MMBR.05019-11
- Dauner, M., Sonderegger, M., Hochuli, M., Szyperski, T., Wüthrich, K., Hohmann, H.-P., Sauer, U., Bailey, J.E., 2002. Intracellular carbon fluxes in riboflavin-producing *Bacillus subtilis* during growth on two-carbon substrate mixtures. *Appl. Environ. Microbiol.* 68, 1760–71.
- Dauner, M., Storni, T., Sauer, U., Sauer, U.W.E., 2001. *Bacillus subtilis* Metabolism and Energetics in Carbon-Limited and Excess-Carbon Chemostat Culture *Bacillus subtilis* Metabolism and Energetics in Carbon-Limited and Excess-Carbon Chemostat Culture. *J. Bacteriol.* 183, 7308–7317. doi:10.1128/JB.183.24.7308
- Davis, D.A., Lynch, H.C., Varley, J., 1999. The production of Surfactin in batch culture by *Bacillus subtilis* ATCC 21332 is strongly influenced by the conditions of nitrogen metabolism. *Enzyme Microb. Technol.* 25, 322–329. doi:10.1016/S0141-0229(99)00048-4
- Debarbouille, M., Gardan, R., Arnaud, M., Rapoport, G., 1999. Role of *bkdR*, a transcriptional activator of the *sigL*-dependent isoleucine and valine degradation

- pathway in *Bacillus subtilis*. J. Bacteriol. 181, 2059–66.
- Debarbouille, M., Gardan, R., Arnaud, M., Rapoport, G., 1999. Role of BkdR , a Transcriptional Activator of the SigL-Dependent Isoleucine and Valine Degradation Pathway in *Bacillus subtilis* Role of BkdR , a Transcriptional Activator of the SigL-Dependent Isoleucine and Valine Degradation Pathway in *Bacillus subtilis* 181, 2059–2066.
- Debois, D., Hamze, K., Guérineau, V., Le Caër, J.-P., Holland, I.B., Lopes, P., Ouazzani, J., Séror, S.J., Brunelle, A., Laprévotte, O., 2008. In situ localisation and quantification of surfactins in a *Bacillus subtilis* swarming community by imaging mass spectrometry. Proteomics 8, 3682–91. doi:10.1002/pmic.200701025
- Detsch, C., Stülke, J., 2003. Ammonium utilization in *Bacillus subtilis*: Transport and regulatory functions of NrgA and NrgB. Microbiology 149, 3289–3297. doi:10.1099/mic.0.26512-0
- Duc, L.H., Hong, H.A., Fairweather, N., Ricca, E., Cutting, S.M., 2003. Bacterial spores as vaccine vehicles. Infect. Immun. 71, 2810–8.
- Dufour, S., Deleu, M., Nott, K., Wathélet, B., Thonart, P., Paquot, M., 2005. Hemolytic activity of new linear surfactin analogs in relation to their physico-chemical properties. Biochim. Biophys. Acta 1726, 87–95. doi:10.1016/j.bbagen.2005.06.015
- Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F., Vater, J., 1999. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc. Natl. Acad. Sci. U. S. A. 96, 13294–9. doi:10.1073/PNAS.96.23.13294
- Duitman, E.H., Wyczawski, D., Boven, L.G., Venema, G., Kuipers, O.P., Hamoen, L.W., 2007. Novel methods for genetic transformation of natural *Bacillus subtilis* isolates used to study the regulation of the mycosubtilin and surfactin synthetases. Appl. Environ. Microbiol. 73, 3490–3496. doi:10.1128/AEM.02751-06
- Ebbighausen, H., Weil, B., Krämer, R., 1989. Transport of branched-chain amino acids in *Corynebacterium glutamicum*. Arch. Microbiol. 151, 238–44.
- Eppelmann, K., Doekel, S., Marahiel, M.A., 2001. Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host *Bacillus subtilis*. J. Biol. Chem. 276, 34824–31. doi:10.1074/jbc.M104456200
- Etchegaray, A., Coutte, F., Chataigné, G., Bechet, M., dos Santos, R.H.Z., Leclère, V., Jacques, P., 2016. Production of *Bacillus amyloliquefaciens* OG and its metabolites in renewable media: valorisation for biodiesel production and p-xylene decontamination. Can. J. Microbiol. cjm-2016-0288. doi:10.1139/cjm-2016-0288
- Fahim, S., Dimitrov, K., Gancel, F., Vauchel, P., Jacques, P., Nikov, I., 2012a. Impact of energy supply and oxygen transfer on selective lipopeptide production by *Bacillus*

- subtilis* BBG21. *Bioresour. Technol.* 126, 1–6. doi:10.1016/j.biortech.2012.09.019
- Fahim, S., Dimitrov, K., Gancel, F., Vauchel, P., Jacques, P., Nikov, I., 2012b. Impact of energy supply and oxygen transfer on selective lipopeptide production by *Bacillus subtilis* BBG21. *Bioresour. Technol.* 126, 1–6. doi:10.1016/j.biortech.2012.09.019
- Farmer, W.R., Liao, J.C., 2001. Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*. *Biotechnol. Prog.* 17, 57–61.
- Fassi Fehri, L., Wróblewski, H., Blanchard, A., 2007. Activities of antimicrobial peptides and synergy with enrofloxacin against *Mycoplasma pulmonis*. *Antimicrob. Agents Chemother.* 51, 468–74. doi:10.1128/AAC.01030-06
- Faulkner, D.J., 1998. Marine natural products. *Nat. Prod. Rep.* 15, 113–58.
- Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., Thomas, M.G., n.d. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* 5, 191–211. doi:10.1021/mp700137g
- Fisher, S.H., 1999. Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la différence! *Mol. Microbiol.* 32, 223–32. doi:10.1046/j.1365-2958.1999.01333.x
- Fravel, D.R., 2005. Commercialization and implementation of biocontrol. *Annu. Rev. Phytopathol.* 43, 337–59. doi:10.1146/annurev.phyto.43.032904.092924
- Fry, B., Zhu, T., Domach, M.M., Koepsel, R.R., Phalakornkule, C., Ataii, M.M., 2000. Characterization of growth and acid formation in a *Bacillus subtilis* pyruvate kinase mutant. *Appl. Environ. Microbiol.* 66, 4045–4049. doi:10.1128/AEM.66.9.4045-4049.2000
- Fujita, Y., Freese, E., 1981. Isolation and properties of a *Bacillus subtilis* mutant unable to produce fructose-bisphosphatase. *J. Bacteriol.* 145, 760–767.
- Fujita, Y., Satomura, T., Tojo, S., Hirooka, K., 2014. CcpA-mediated catabolite activation of the *Bacillus subtilis* *ilv-leu* operon and its negation by either CodY- or TnrA-mediated negative regulation. *J. Bacteriol.* 196, 3793–3806. doi:10.1128/JB.02055-14
- Fujita, Y., Yoshida, K., Miwa, Y., Yanai, N., Nagakawa, E., Kasahara, Y., 1998. Identification and expression of the *Bacillus subtilis* fructose-1, 6-bisphosphatase gene (*fbp*). *J. Bacteriol.* 180, 4309–4313.
- Gao, S., Wu, H., Yu, X., Qian, L., Gao, X., 2016. Swarming motility plays the major role in migration during tomato root colonization by *Bacillus subtilis* SWR01. *Biol. Control* 98, 11–17. doi:10.1016/j.biocontrol.2016.03.011
- Gaur, N.K., Dubnau, E., Smith, I., 1986. Characterization of a cloned *Bacillus subtilis* gene that inhibits sporulation in multiple copies. *J. Bacteriol.* 168, 860–9.
- Gaur, N.K., Oppenheim, J., Smith, I., 1991. The *Bacillus subtilis* *sin* gene, a regulator of alternate developmental processes, codes for a DNA-binding protein. *J. Bacteriol.*

173, 678–86.

- Gay, N.J., Gangloff, M., 2007. Structure and function of Toll receptors and their ligands. *Annu. Rev. Biochem.* 76, 141–65. doi:10.1146/annurev.biochem.76.060305.151318
- Ghojavand, H., Vahabzadeh, F., Roayaei, E., Shahraki, A.K., 2008. Production and properties of a biosurfactant obtained from a member of the *Bacillus subtilis* group (PTCC 1696). *J. Colloid Interface Sci.* 324, 172–6. doi:10.1016/j.jcis.2008.05.001
- Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessières, P., Aymerich, S., Fromion, V., 2008. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst. Biol.* 2, 20. doi:10.1186/1752-0509-2-20
- Gong, G., Zheng, Z., Chen, H., Yuan, C., Wang, P., Yao, L., Yu, Z., n.d. Enhanced Production of Surfactin by *Bacillus subtilis* E8 Mutant Obtained by Ion Beam Implantation.
- Grandoni, J.A., Zahler, S.A., Calvo, J.M., 1992. Transcriptional regulation of the *ilv-leu* operon of *Bacillus subtilis*. *J. Bacteriol.* 174, 3212–3219.
- Grau, A., Gómez Fernández, J.C., Peypoux, F., Ortiz, A., 1999. A study on the interactions of surfactin with phospholipid vesicles. *Biochim. Biophys. Acta* 1418, 307–19.
- Green, B.D., Olmedo, G., Youngman, P., n.d. A genetic analysis of Spo0A structure and function. *Res. Microbiol.* 142, 825–30.
- Guan, C., Cui, W., Cheng, J., Zhou, L., Guo, J., Hu, X., Xiao, G., Zhou, Z., 2015. Construction and development of an auto-regulatory gene expression system in *Bacillus subtilis*. *Microb. Cell Fact.* 14, 150. doi:10.1186/s12934-015-0341-2
- Guan, C., Cui, W., Cheng, J., Zhou, L., Liu, Z., Zhou, Z., 2016. Development of an efficient autoinducible expression system by promoter engineering in *Bacillus subtilis*. *Microb. Cell Fact.* 1–12. doi:10.1186/s12934-016-0464-0
- Guez, J.S., Muller, C.H., Danze, P.M., B??chs, J., Jacques, P., 2008. Respiration activity monitoring system (RAMOS), an efficient tool to study the influence of the oxygen transfer rate on the synthesis of lipopeptide by *Bacillus subtilis* ATCC6633. *J. Biotechnol.* 134, 121–126. doi:10.1016/j.jbiotec.2008.01.003
- Gupta, R., Beg, Q.K., Lorenz, P., 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* 59, 15–32. doi:10.1007/s00253-002-0975-y
- Haese, A., Schubert, M., Herrmann, M., Zocher, R., 1993. Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. *Mol. Microbiol.* 7, 905–14.
- Hahn, J., Dubnau, D., 1991. Growth stage signal transduction and the requirements for *srfA* induction in development of competence. *J. Bacteriol.* 173, 7275–82.

- Hahn, J., Roggiani, M., Dubnau, D., 1995. The major role of Spo0A in genetic competence is to downregulate *abrB*, an essential competence gene. *J. Bacteriol.* 177, 3601–5.
- Haijema, B.J., van Sinderen, D., Winterling, K., Kooistra, J., Venema, G., Hamoen, L.W., 1996. Regulated expression of the *dinR* and *recA* genes during competence development and SOS induction in *Bacillus subtilis*. *Mol. Microbiol.* 22, 75–85.
- Hamilton, C.M., Aldea, M., Washburn, B.K., Babitzke, P., Kushner, S.R., 1989. New method for generating deletions and gene replacements in *Escherichia coli*. *J. Bacteriol.* 171, 4617–22.
- Hamoen, L.W., Eshuis, H., Jongbloed, J., Venema, G., van Sinderen, D., 1995. A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Mol. Microbiol.* 15, 55–63.
- Hamoen, L.W., Venema, G., Kuipers, O.P., 2003. Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* 149, 9–17. doi:10.1099/mic.0.26003-0
- Handke, L.D., Shivers, R.P., Sonenshein, A.L., 2008. Interaction of *Bacillus subtilis* CodY with GTP. *J. Bacteriol.* 190, 798–806. doi:10.1128/JB.01115-07
- Hatfield, G.W., Umbarger, H.E., 1970. Threonine deaminase from *Bacillus subtilis*. II. The steady state kinetic properties. *J. Biol. Chem.* 245, 1742–7.
- Hbid, C., Jacques, P., Razafindralambo, H., Mpoyo, M.K., Meurice, E., Paquot, M., Thonart, P., 1996. Influence of the production of two lipopeptides, iturin a and surfactin s1, on oxygen transfer during *Bacillus subtilis* fermentation. *Appl Biochem Biotechnol* 57–8, 571–579. doi:10.1007/BF02941737
- Heerklotz, H., Seelig, J., 2001. Detergent-like action of the antibiotic peptide surfactin on lipid membranes. *Biophys. J.* 81, 1547–54. doi:10.1016/S0006-3495(01)75808-0
- Hénaut, A., Rouxel, T., Gleizes, A., Moszer, I., Danchin, A., 1996. Uneven Distribution of GATC Motifs in the *Escherichia coli* Chromosome, its Plasmids and its Phages. *J. Mol. Biol.* 257, 574–585. doi:10.1006/jmbi.1996.0186
- Henry, G., Deleu, M., Jourdan, E., Thonart, P., Ongena, M., 2011. The bacterial lipopeptide surfactin targets the lipid fraction of the plant plasma membrane to trigger immune-related defence responses. *Cell. Microbiol.* 13, 1824–1837. doi:10.1111/j.1462-5822.2011.01664.x
- Herbort, M., Klein, M., Manting, E.H., Driessen, A.J., Freudl, R., 1999. Temporal expression of the *Bacillus subtilis* *secA* gene, encoding a central component of the preprotein translocase. *J. Bacteriol.* 181, 493–500.
- Hoffmann, K., Schneider-Scherzer, E., Kleinkauf, H., Zocher, R., 1994. Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *J. Biol. Chem.* 269, 12710–4.

- HOSONO, K., SUZUKI, H., 1983. Acylpeptides, the inhibitors of cyclic adenosine 3',5'-monophosphate phosphodiesterase. I. Purification, physicochemical properties and structures of fatty acid residues. *J. Antibiot. (Tokyo)*. 36, 667–673. doi:10.7164/antibiotics.36.667
- Hourdou, M.L., Besson, F., Tenoux, I., Michel, G., 1989. Fatty acid and beta-amino acid syntheses in strains of *Bacillus subtilis* producing iturinic antibiotics. *Lipids* 24, 940–4.
- Huang, X., Liu, J., Wang, Y., Liu, J., Lu, L., 2015. The positive effects of Mn²⁺ on nitrogen use and surfactin production by *Bacillus subtilis* ATCC 21332. *Biotechnol. Biotechnol. Equip.* 29, 381–389. doi:10.1080/13102818.2015.1006905
- Hwang, M.-H., Lim, J.-H., Yun, H.-I., Rhee, M.-H., Cho, J.-Y., Hsu, W.H., Park, S.-C., 2005. Surfactin C inhibits the lipopolysaccharide-induced transcription of interleukin-1beta and inducible nitric oxide synthase and nitric oxide production in murine RAW 264.7 cells. *Biotechnol. Lett.* 27, 1605–8. doi:10.1007/s10529-005-2515-1
- Hwang, Y.-H., Park, B.-K., Lim, J.-H., Kim, M.-S., Park, S.-C., Hwang, M.-H., Yun, H.-I., 2007. Lipopolysaccharide-binding and neutralizing activities of surfactin C in experimental models of septic shock. *Eur. J. Pharmacol.* 556, 166–71. doi:10.1016/j.ejphar.2006.10.031
- Jacques, P., 2011. Biosurfactants. doi:10.1007/978-3-642-14490-5
- Jiang, J., Gao, L., Bie, X., Lu, Z., Liu, H., Zhang, C., Lu, F., Zhao, H., 2016. Identification of novel surfactin derivatives from NRPS modification of *Bacillus subtilis* and its antifungal activity against *Fusarium moniliforme*. *BMC Microbiol.* 16, 31. doi:10.1186/s12866-016-0645-3
- Jourdan, E., Henry, G., Duby, F., Dommes, J., Barthélemy, J.P., Thonart, P., Ongena, M., 2009. Insights into the defense-related events occurring in plant cells following perception of surfactin-type lipopeptide from *Bacillus subtilis*. *Mol. Plant. Microbe Interact.* 22, 456–68. doi:10.1094/MPMI-22-4-0456
- Jules, M., Chat, L.L., Aymerich, S., Coq, D.L., 2009. The *Bacillus subtilis* *ywjI* (*glpX*) gene encodes a class II fructose-1,6-bisphosphatase, functionally equivalent to the class III Fbp enzyme. *J. Bacteriol.* 191, 3168–3171. doi:10.1128/JB.01783-08
- Kakinuma, A., Tamura, G., Arima, K., 1968. Wetting of fibrin plate and apparent promotion of fibrinolysis by surfactin, a new bacterial peptidelipid surfactant. *Experientia* 24, 1120–1.
- Kameda, Y., Oira, S., Matsui, K., Kanatomo, S., Hase, T., 1974. Antitumor activity of bacillus natto. V. Isolation and characterization of surfactin in the culture medium of *Bacillus natto* KMD 2311. *Chem. Pharm. Bull. (Tokyo)*. 22, 938–44.
- Kaneda, T., 1991. Iso- and Anteiso-Fatty Acids in Bacteria: Biosynthesis, Function, and Taxonomic Significance. *Microbiol. Rev.* 55, 288–302.

- Kaneda, T., 1991. Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol. Rev.* 55, 288–302.
- Kaneda, T., 1977. Fatty acids of the genus *Bacillus*: an example of branched-chain preference. *Bacteriol. Rev.* 41, 391–418.
- Karl Lintner, 2004. US20040132667.pdf.
- Kealey, J.T., Liu, L., Santi, D. V, Betlach, M.C., Barr, P.J., 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc. Natl. Acad. Sci. U. S. A.* 95, 505–9.
- Kearns, D.B., 2010. A field guide to bacterial swarming motility. *Nat. Rev. Microbiol.* 8, 634–44. doi:10.1038/nrmicro2405
- Kearns, D.B., Chu, F., Rudner, R., Losick, R., 2004. Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. *Mol. Microbiol.* 52, 357–369. doi:10.1111/j.1365-2958.2004.03996.x
- Keating, T.A., Ehmann, D.E., Kohli, R.M., Marshall, C.G., Trauger, J.W., Walsh, C.T., 2001. Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. *ChemBiochem* 2, 99–107.
- Kei Arima, Gakuzo Tamura, A.K., 1972. Surfactin.
- Kennerknecht, N., Sahm, H., Yen, M.-R., Pátek, M., Saier Jr, M.H., Eggeling, L., 2002. Export of L-isoleucine from *Corynebacterium glutamicum*: a two-gene-encoded member of a new translocator family. *J. Bacteriol.* 184, 3947–56.
- Kikuchi, T., Hasumi, K., 2002. Enhancement of plasminogen activation by surfactin C: augmentation of fibrinolysis in vitro and in vivo. *Biochim. Biophys. Acta* 1596, 234–45.
- Kim, H., Yoon, B., Lee, C., Katsuragi, T., Tani, Y., Oh, H., 1997. Production and Properties of a Lipopeptide Biosurfactant from *Bacillus subtilis* C9.
- Kim, S.-Y., Kim, J.Y., Kim, S.-H., Bae, H.J., Yi, H., Yoon, S.H., Koo, B.S., Kwon, M., Cho, J.Y., Lee, C.-E., Hong, S., 2007. Surfactin from *Bacillus subtilis* displays anti-proliferative effect via apoptosis induction, cell cycle arrest and survival signaling suppression. *FEBS Lett.* 581, 865–71. doi:10.1016/j.febslet.2007.01.059
- Kim, S.B., Shin, B.S., Choi, S.K., Kim, C.K., Park, S.H., 2001. Involvement of acetyl phosphate in the in vivo activation of the response regulator ComA in *Bacillus subtilis*. *FEMS Microbiol. Lett.* 195, 179–83.
- Kim, S.D., Park, S.K., Cho, J.Y., Park, H.J., Lim, J.H., Yun, H.I., Park, S.C., Lee, K.Y., Kim, S.K., Rhee, M.H., 2006. Surfactin C inhibits platelet aggregation. *J. Pharm. Pharmacol.* 58, 867–70. doi:10.1211/jpp.58.6.0018
- Klein, W., Weber, M.H., Marahiel, M.A., 1999. Cold shock response of *Bacillus subtilis*:

- isoleucine-dependent switch in the fatty acid branching pattern for membrane adaptation to low temperatures. *J. Bacteriol.* 181, 5341–9.
- Koglin, A., Mofid, M.R., Löhr, F., Schäfer, B., Rogov, V. V, Blum, M.-M., Mittag, T., Marahiel, M.A., Bernhard, F., Dötsch, V., 2006. Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* 312, 273–6. doi:10.1126/science.1122928
- Kohlstedt, M., Sappa, P.K., Meyer, H., Maaß, S., Zaprasia, A., Hoffmann, T., Becker, J., Steil, L., Hecker, M., van Dijl, J.M., aarten, Lalk, M., Mader, U., Stulke, J., Bremer, E., Volker, U., Wittmann, C., 2014a. Adaptation of *Bacillus subtilis* carbon core metabolism to simultaneous nutrient limitation and osmotic challenge: a multi-omics perspective. *Environ. Microbiol.* 16, 1898–1917. doi:10.1111/1462-2920.12438
- Kohlstedt, M., Sappa, P.K., Meyer, H., Maaß, S., Zaprasia, A., Hoffmann, T., Becker, J., Steil, L., Hecker, M., van Dijl, J.M., Lalk, M., Mäder, U., Stülke, J., Bremer, E., Völker, U., Wittmann, C., 2014b. Adaptation of *Bacillus subtilis* carbon core metabolism to simultaneous nutrient limitation and osmotic challenge: A multi-omics perspective. *Environ. Microbiol.* 16, 1898–1917. doi:10.1111/1462-2920.12438
- Kopp, F., Marahiel, M.A., 2007. Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. *Curr. Opin. Biotechnol.* 18, 513–20. doi:10.1016/j.copbio.2007.09.009
- Kowall, M., Vater, J., Kluge, B., Stein, T., Franke, P., Ziessow, D., 1998. Separation and Characterization of Surfactin Isoforms Produced by *Bacillus subtilis* OKB 105. *J. Colloid Interface Sci.* 204, 1–8. doi:10.1006/jcis.1998.5558
- Kracht, M., Rokos, H., Ozel, M., Kowall, M., Pauli, G., Vater, J., 1999. Antiviral and hemolytic activities of surfactin isoforms and their methyl ester derivatives. *J. Antibiot. (Tokyo)*. 52, 613–619.
- Kunst, F., al., et, 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390, 249–256.
- Lai, J.R., Fischbach, M.A., Liu, D.R., Walsh, C.T., 2006. Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. *J. Am. Chem. Soc.* 128, 11002–3. doi:10.1021/ja063238h
- Lam, K.H.E., Chow, K.C., Wong, W.K.R., 1998. Construction of an efficient *Bacillus subtilis* system for extracellular production of heterologous proteins. *J. Biotechnol.* 63, 167–177. doi:10.1016/S0168-1656(98)00041-8
- Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., Walsh, C.T., 1996. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* 3, 923–36.
- LANDY, M., WARREN, G.H., 1948. Bacillomycin; an antibiotic from *Bacillus subtilis* active against pathogenic fungi. *Proc. Soc. Exp. Biol. Med.* 67, 539–41.

- Lazazzera, B.A., Kurtser, I.G., McQuade, R.S., Grossman, A.D., 1999. An autoregulatory circuit affecting peptide signaling in *Bacillus subtilis*. *J. Bacteriol.* 181, 5193–200.
- Lazazzera, B.A., Solomon, J.M., Grossman, A.D., 1997. An exported peptide functions intracellularly to contribute to cell density signaling in *Bacillus subtilis*. *Cell* 89, 917–25.
- Leclère, V., Béchet, M., Adam, A., Guez, J.-S., Wathelet, B., Ongena, M., Thonart, P., Gancel, F., Chollet-Imbert, M., Jacques, P., 2005. Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. *Appl. Environ. Microbiol.* 71, 4577–84. doi:10.1128/AEM.71.8.4577-4584.2005
- Leenhouts, K., Buist, G., Bolhuis, A., Ten Berge, A., Kiel, J., Mierau, I., Dabrowska, M., Venema, G., Kok, J., 1996. A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol. Gen. Genet.* 253, 217–224. doi:10.1007/s004380050315
- Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C., Stülke, J., 2010. The RNA degradosome in *Bacillus subtilis*: Identification of *csha* as the major RNA helicase in the multiprotein complex. *Mol. Microbiol.* 77, 958–971. doi:10.1111/j.1365-2958.2010.07264.x
- Li, S., Huang, D., Li, Y., Wen, J., Jia, X., 2012. Rational improvement of the engineered isobutanol-producing *Bacillus subtilis* by elementary mode analysis. *Microb. Cell Fact.* 11, 101. doi:10.1186/1475-2859-11-101
- Li, X., Yang, H., Zhang, D., Li, X., Yu, H., Shen, Z., 2015. Overexpression of specific proton motive force-dependent transporters facilitate the export of surfactin in *Bacillus subtilis*. *J. Ind. Microbiol. Biotechnol.* 42, 93–103. doi:10.1007/s10295-014-1527-z
- Licona-Cassani, C., Lara, A.R., Cabrera-Valladares, N., Escalante, A., Hernández-Chávez, G., Martínez, A., Bolívar, F., Gosset, G., 2014. Inactivation of pyruvate kinase or the phosphoenolpyruvate: Sugar phosphotransferase system increases shikimic and dehydroshikimic acid yields from glucose in *Bacillus subtilis*. *J. Mol. Microbiol. Biotechnol.* 24, 37–45. doi:10.1159/000355264
- Lin, J.Y., Prasad, C., 1974. Selection of a mutant of *Bacillus subtilis* deficient in glucose-6-phosphate dehydrogenase and phosphoglucoisomerase. *J. Gen. Microbiol.* 83, 419–421. doi:10.1099/00221287-83-2-419
- Lipmann, F., Gevers, W., Kleinkauf, H., Roskoski, R., 1971. Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. *Adv. Enzymol. Relat. Areas Mol. Biol.* 35, 1–34.
- Liu, J.F., Yang, J., Yang, S.Z., Ye, R.Q., Mu, B.Z., 2012. Effects of different amino acids in culture media on surfactin variants produced by *Bacillus subtilis* TD7. *Appl. Biochem. Biotechnol.* 166, 2091–2100. doi:10.1007/s12010-012-9636-5

- Liu, X., Ren, B., Gao, H., Liu, M., Dai, H., Song, F., Yu, Z., Wang, S., Hu, J., Kokare, C.R., Zhang, L., 2012. Optimization for the production of surfactin with a new synergistic antifungal activity. *PLoS One* 7, e34430. doi:10.1371/journal.pone.0034430
- Liu, X.Y., Yang, S.Z., Mu, B.Z., 2009. Production and characterization of a C15-surfactin-O-methyl ester by a lipopeptide producing strain *Bacillus subtilis* HSO121. *Process Biochem.* 44, 1144–1151. doi:10.1016/j.procbio.2009.06.014
- Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F.M., Hecker, M., Stülke, J., 2001. Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the *gapA* operon. *Mol. Microbiol.* 41, 409–22.
- Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F.M., Hecker, M., Stülke, J., 2001. Transcription of glycolytic genes and operons in *Bacillus subtilis*: Evidence for the presence of multiple levels of control of the *gapA* operon. *Mol. Microbiol.* 41, 409–422. doi:10.1046/j.1365-2958.2001.02523.x
- Macek, B., Mijakovic, I., Olsen, J. V, Gnad, F., Kumar, C., Jensen, P.R., Mann, M., 2007. The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol. Cell. Proteomics* 6, 697–707. doi:10.1074/mcp.M600464-MCP200
- Mäder, U., Hennig, S., Hecker, M., Homuth, G., 2004a. Transcriptional organization and posttranscriptional regulation of the *Bacillus subtilis* branched-chain amino acid biosynthesis genes. *J. Bacteriol.* 186, 2240–52.
- Mäder, U., Hennig, S., Hecker, M., Homuth, G., 2004b. Transcriptional Organization and Posttranscriptional Regulation of the *Bacillus subtilis* Branched-Chain Amino Acid Biosynthesis Genes 186, 2240–2252. doi:10.1128/JB.186.8.2240
- Mäder, U., Schmeisky, A.G., Flórez, L.A., Stülke, J., 2012. SubtiWiki--a comprehensive community resource for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* 40, D1278-87. doi:10.1093/nar/gkr923
- Maget-Dana, R., Ptak, M., 1995. Interactions of surfactin with membrane models. *Biophys. J.* 68, 1937–43. doi:10.1016/S0006-3495(95)80370-X
- Maget-Dana, R., Ptak, M., 1992. Interfacial properties of surfactin. *J. Colloid Interface Sci.* 153, 285–291. doi:10.1016/0021-9797(92)90319-H
- Marahiel, M.A., 2009. Working outside the protein-synthesis rules: Insights into non-ribosomal peptide synthesis. *J. Pept. Sci.* 15, 799–807. doi:10.1002/psc.1183
- Menkhaus, M., Ullrich, C., Kluge, B., Vater, J., Vollenbroich, D., Kamp, R.M., 1993. Structural and functional organization of the surfactin synthetase multienzyme system. *J. Biol. Chem.* 268, 7678–84.
- Merino, E., Yanofsky, C., Winkler, W., al., et, Henkin, T.M., Yanofsky, C., Yanofsky, C., Vitreschak, A.G., al., et, Merino, E., Yanofsky, C., Vitreschak, A.G., al., et, Barrick,

- J.E., al., et, Abreu-Goodger, C., al., et, Moreno-Hagelsieb, G., Collado-Vides, J., Tatusov, R.L., al., et, Yanofsky, C., Grundy, F.J., Henkin, T.M., Vitreschak, A.G., al., et, McDaniel, B.A., al., et, Gollnick, P., al., et, Amster-Choder, O., Wright, A., Zuker, M., Stiegler, P., 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* 21, 260–4. doi:10.1016/j.tig.2005.03.002
- Meyer, F.M., Stülke, J., 2013. Malate metabolism in *Bacillus subtilis*: Distinct roles for three classes of malate-oxidizing enzymes. *FEMS Microbiol. Lett.* 339, 17–22. doi:10.1111/1574-6968.12041
- Mireles, J.R., Toguchi, A., Harshey, R.M., 2001. *Salmonella enterica* serovar typhimurium swarming mutants with altered biofilm-forming abilities: surfactin inhibits biofilm formation. *J. Bacteriol.* 183, 5848–54. doi:10.1128/JB.183.20.5848-5854.2001
- Mofid, M.R., Finking, R., Essen, L.O., Marahiel, M.A., 2004. Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition and reaction mechanism. *Biochemistry* 43, 4128–36. doi:10.1021/bi036013h
- Molle, V., Nakaura, Y., Shivers, R.P., Yamaguchi, H., Losick, R., Fujita, Y., Sonenshein, A.L., 2003. Additional targets of the *Bacillus subtilis* global regulator codY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J. Bacteriol.* 185, 1911–1922. doi:10.1128/JB.185.6.1911-1922.2003
- Monahan, L.G., Hajduk, I. V., Blaber, S.P., Charles, I.G., Harry, E.J., 2014. Coordinating bacterial cell division with nutrient availability: A role for glycolysis. *MBio* 5, 1–13. doi:10.1128/mBio.00935-14
- Mootz, H.D., Finking, R., Marahiel, M.A., 2001. 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J. Biol. Chem.* 276, 37289–98. doi:10.1074/jbc.M103556200
- Mootz, H.D., Marahiel, M.A., 1997. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* 179, 6843–50.
- Mootz, H.D., Schwarzer, D., Marahiel, M.A., 2002. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBiochem* 3, 490–504. doi:10.1002/1439-7633(20020603)3:6490::AID-CBIC4903.0.CO;2-N
- Motta Dos Santos, L.F., Coutte, F., Ravallec, R., Dhulster, P., Tournier-Couturier, L., Jacques, P., 2016. An improvement of surfactin production by *Bacillus subtilis* BBG131 using design of experiments in microbioreactors and continuous process in bubbleless membrane bioreactor. *Bioresour. Technol.* 218, 944–52. doi:10.1016/j.biortech.2016.07.053
- Mulligan, C.N., 2005. Environmental applications for biosurfactants. *Environ. Pollut.* 133, 183–98. doi:10.1016/j.envpol.2004.06.009
- Muñoz-Márquez, M.E., Ponce-Rivas, E., 2010. Effect of *pfkA* chromosomal interruption

- on growth, sporulation, and production of organic acids in *Bacillus subtilis*. J. Basic Microbiol. 50, 232–240. doi:10.1002/jobm.200900236
- Naik, D.N., Kaneda, T., 1974. Biosynthesis of branched long-chain fatty acids by species of *Bacillus*: relative activity of three alpha-keto acid substrates and factors affecting chain length. Can. J. Microbiol. 20, 1701–8.
- Nakano, M.M., Corbell, N., Besson, J., Zuber, P., 1992. Isolation and characterization of *sfp*: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. Mol. Gen. Genet. MGG 232, 313–321. doi:10.1007/bf00280011
- Nakano, M.M., Marahiel, M.A., Zuber, P., 1988. Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. J. Bacteriol. 170, 5662–5668.
- Nakano, M.M., Xia, L.A., Zuber, P., 1991. Transcription initiation region of the *srfA* operon, which is controlled by the comP-comA signal transduction system in *Bacillus subtilis*. J. Bacteriol. 173, 5487–93.
- Newman, J.A., Hewitt, L., Rodrigues, C., Solovyova, A.S., Harwood, C.R., Lewis, R.J., 2012. Dissection of the network of interactions that links RNA processing with glycolysis in the *Bacillus subtilis* degradosome. J. Mol. Biol. 416, 121–136. doi:10.1016/j.jmb.2011.12.024
- Nickel, M., Homuth, G., Böhnisch, C., Mäder, U., Schweder, T., 2004. Cold induction of the *Bacillus subtilis* bkd operon is mediated by increased mRNA stability. Mol. Genet. Genomics 272, 98–107. doi:10.1007/s00438-004-1038-0
- Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., Becher, D., Bisicchia, P., Botella, E., Delumeau, O., Doherty, G., Denham, E.L., Fogg, M.J., Fromion, V., Goelzer, A., Hansen, A., Härtig, E., Harwood, C.R., Homuth, G., Jarmer, H., 2012. Condition-Dependent Transcriptome Architecture in *Bacillus subtilis*. Science (80-.). 1103, 1103–1106. doi:10.1126/science.1206848
- Niehren, J., Versari, C., John, M., Coutte, F., Jacques, P., 2016. Predicting Changes of Reaction Networks with Partial Kinetic Information. BioSystems 47.
- Noudeh, G.D., Noodeh, A.D., Moshafi, M.H., Behravan, E., Afzadi, M.A., Sodagar, M., 2010. Investigation of cellular hydrophobicity and surface activity effects of biosynthesized biosurfactant from broth media of PTCC 1561. African J. Microbiol. Res. 4, 1814–1822.
- Olano, C., Lombó, F., Méndez, C., Salas, J.A., 2008. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. Metab. Eng. 10, 281–292. doi:10.1016/j.ymben.2008.07.001
- Ongena, M., Jacques, P., 2008. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol. 16, 115–25. doi:10.1016/j.tim.2007.12.009

- Otero, J.M., Nielsen, J., 2010. Industrial systems biology. *Biotechnol. Bioeng.* 105, 439–60. doi:10.1002/bit.22592
- Pandey, A., Nigam, P., Soccol, C.R., Soccol, V.T., Singh, D., Mohan, R., 2000. Advances in microbial amylases. *Biotechnol. Appl. Biochem.* 31 (Pt 2), 135–52.
- Paolo Carrera, Paola Cosmina, G.G., 1993. Mutant of *Bacillus subtilis*.
- Perego, M., Glaser, P., Hoch, J.A., 1996. Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. *Mol. Microbiol.* 19, 1151–7.
- Perego, M., Hoch, J.A., 1988. Sequence analysis and regulation of the hpr locus, a regulatory gene for protease production and sporulation in *Bacillus subtilis*. *J. Bacteriol.* 170, 2560–7.
- Peypoux, F., Bonmatin, J.M., Wallach, J., 1999c. Recent trends in the biochemistry of surfactin. *Appl. Microbiol. Biotechnol.* 51, 553–563.
- Phillips, Z.E. V, Strauch, M.A., 2002. *Bacillus subtilis* sporulation and stationary phase gene expression. *Cell. Mol. Life Sci.* 59, 392–402.
- Plaza, G., Chojniak, J., Rudnicka, K., Paraszkiwicz, K., Bernat, P., 2015. Detection of biosurfactants in *Bacillus* species: Genes and products identification. *J. Appl. Microbiol.* 119, 1023–1034. doi:10.1111/jam.12893
- Raaijmakers, J.M., De Bruijn, I., Nybroe, O., Ongena, M., 2010. Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol. Rev.* 34, 1037–62. doi:10.1111/j.1574-6976.2010.00221.x
- Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Dürr, F., Emenegger, J., 2013. The *Bacillus* BioBrick Box : generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis* The *Bacillus* BioBrick Box : generation and evaluation of essential genetic building blocks for standardized work with B.
- Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 6. doi:10.1371/journal.pcbi.1000744
- Ratnayake-Lecamwasam, M., Serror, P., Wong, K.W., Sonenshein, A.L., 2001. *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Genes Dev.* 15, 1093–103. doi:10.1101/gad.874201
- Reuß, D.R., Thürmer, A., Daniel, R., Quax, W.J., Stülke, J., 2016. Complete Genome Sequence of *Bacillus subtilis* subsp. *subtilis* Strain $\Delta 6$. *Genome Announc.* 4. doi:10.1128/genomeA.00759-16
- Rollins, S.M., 2014. Induced Transcriptional Expression of *Bacillus subtilis* Amino Acid Permease yvbW in Response to Leucine Limitation 484–492.
- Romero, S., Merino, E., Bolívar, F., Gosset, G., Martinez, A., 2007. Metabolic engineering

- of *Bacillus subtilis* for ethanol production: Lactate dehydrogenase plays a key role in fermentative metabolism. *Appl. Environ. Microbiol.* 73, 5190–5198.
doi:10.1128/AEM.00625-07
- S A Sieber, M. a M., 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* 105, 715–738.
doi:10.1021/cr0301191
- Sajitha, K.L., Dev, S.A., Maria Florence, E.J., 2016. Identification and Characterization of Lipopeptides from *Bacillus subtilis* B1 Against Sapstain Fungus of Rubberwood Through MALDI-TOF-MS and RT-PCR. *Curr. Microbiol.* 73, 46–53.
doi:10.1007/s00284-016-1025-9
- Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., Essen, L.-O., 2007. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 15, 781–92. doi:10.1016/j.str.2007.05.008
- Samel, S.A., Wagner, B., Marahiel, M.A., Essen, L.-O., 2006. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J. Mol. Biol.* 359, 876–89. doi:10.1016/j.jmb.2006.03.062
- Schaller, K.D., Fox, S.L., Bruhn, D.F., Noah, K.S., Bala, G.A., 2004. Characterization of surfactin from *Bacillus subtilis* for application as an agent for enhanced oil recovery. *Appl. Biochem. Biotechnol.* 113–116, 827–36.
- Schauwecker, F., Pfennig, F., Grammel, N., Keller, U., 2000. Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. *Chem. Biol.* 7, 287–97.
- Schwarzer, D., Mootz, H.D., Linne, U., Marahiel, M.A., 2002. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14083–8. doi:10.1073/pnas.212382199
- Serror, P., Sonenshein, A.L., 1996a. CodY is required for nutritional repression of *Bacillus subtilis* genetic competence. *J. Bacteriol.* 178, 5910–5.
- Servant, P., Le Coq, D., Aymerich, S., 2005. CcpN (YqzB), a novel regulator for CcpA-independent catabolite repression of *Bacillus subtilis* gluconeogenic genes. *Mol. Microbiol.* 55, 1435–1451. doi:10.1111/j.1365-2958.2005.04473.x
- Seydlová, G., Svobodová, J., 2008. Review of surfactin chemical properties and the potential biomedical applications. *Open Med.* 3, 123–133. doi:10.2478/s11536-008-0002-5
- Seydlová, G., Svobodová, J., 2008. Development of membrane lipids in the surfactin producer *Bacillus subtilis*. *Folia Microbiol. (Praha).* 53, 303–307.
doi:10.1007/s12223-008-0047-5
- Shaligram, N.S., Singhal, R.S., 2010. Surfactin -a review on biosynthesis, fermentation, purification and applications. *Food Technol. Biotechnol.* 48, 119–134.

- Shao, C., Liu, L., Gang, H., Yang, S., Mu, B., 2015. Structural diversity of the microbial surfactin derivatives from selective esterification approach. *Int. J. Mol. Sci.* 16, 1855–1872. doi:10.3390/ijms16011855
- Sheppard, J., Cooper, D., 1991. The response of *Bacillus subtilis* ATCC 21332 to manganese during continuous-phased growth. *Appl. Microbiol. Biotechnol.* 35, 72–76. doi:10.1007/BF00180639
- Shivers, R.P., Sonenshein, A.L., 2004. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol. Microbiol.* 53, 599–611. doi:10.1111/j.1365-2958.2004.04135.x
- Sieber, S.A., Marahiel, M.A., 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* 105, 715–38. doi:10.1021/cr0301191
- Sohn, S.B., Graf, A.B., Kim, T.Y., Gasser, B., Maurer, M., Ferrer, P., Mattanovich, D., Lee, S.Y., 2010. Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for *in silico* analysis of heterologous protein production. *Biotechnol. J.* 5, 705–715. doi:10.1002/biot.201000078
- Solomon, J.M., Lazazzera, B.A., Grossman, A.D., 1996. Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes Dev.* 10, 2014–24.
- Solomon, J.M., Magnuson, R., Srivastava, A., Grossman, A.D., 1995. Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes Dev.* 9, 547–58.
- Sonenshein, A.L., 2007. Control of key metabolic intersections in *Bacillus subtilis*. *Nat. Rev. Microbiol.* 5, 917–27. doi:10.1038/nrmicro1772
- Stachelhaus, T., Mootz, H.D., Bergendahl, V., Marahiel, M.A., 1998. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* 273, 22773–81.
- Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 6, 493–505. doi:10.1016/S1074-5521(99)80082-9
- Stanley, N.R., Lazazzera, B.A., 2004. Environmental signals and regulatory pathways that influence biofilm formation. *Mol. Microbiol.* 52, 917–24. doi:10.1111/j.1365-2958.2004.04036.x
- Steller, S., Sokoll, A., Wilde, C., Bernhard, F., Franke, P., Vater, J., 2004. Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* 43, 11331–43. doi:10.1021/bi0493416
- Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., Jacques, P., Thonart, P., Vater, J., 1999. Structural and functional organization of the fengycin

synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chem. Biol.* 6, 31–41. doi:10.1016/S1074-5521(99)80018-0

- Strauch, M.A., Wu, J.J., Jonas, R.H., Hoch, J.A., 1993. A positive feedback loop controls transcription of the *spoOF* gene, a component of the sporulation phosphorelay in *Bacillus subtilis*. *Mol. Microbiol.* 7, 967–74.
- Sun, H., Bie, X., Lu, F., Lu, Y., Wu, Y., Lu, Z., 2009. NOTE / NOTE Enhancement of surfactin production of *Bacillus subtilis* *fmbR* by replacement of the native promoter with the Pspac promoter 1006, 1003–1006. doi:10.1139/W09-044
- Tadashi Yoneda, Yoshiaki Miyota, Kazuo Furuya, T.T., 2006. Production process of surfactin.
- Takahashi, T., Ohno, O., Ikeda, Y., Sawa, R., Homma, Y., Igarashi, M., Umezawa, K., 2006. Inhibition of lipopolysaccharide activity by a bacterial cyclic lipopeptide surfactin. *J. Antibiot. (Tokyo)*. 59, 35–43. doi:10.1038/ja.2006.6
- Tanaka, K., Henry, C.S., Zinner, J.F., Jolivet, E., Cohoon, M.P., Xia, F., Bidnenko, V., Ehrlich, S.D., Stevens, R.L., Noiro, P., 2013. Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res.* 41, 687–699. doi:10.1093/nar/gks963
- Tang, J.S., Zhao, F., Gao, H., Dai, Y., Yao, Z.H., Hong, K., Li, J., Ye, W.C., Yao, X.S., 2010. Characterization and online detection of surfactin isomers based on HPLC-MS analyses and their inhibitory effects on the overproduction of nitric oxide and the release of TNF and IL-6 in LPS-induced macrophages. *Mar. Drugs* 8, 2605–2618. doi:10.3390/md8102605
- Thomaidis, H.B., Davison, E.J., Burston, L., Johnson, H., Brown, D.R., Hunt, A.C., Errington, J., Czaplowski, L., 2007. Essential bacterial functions encoded by gene pairs. *J. Bacteriol.* 189, 591–602. doi:10.1128/JB.01381-06
- Tobisch, S., Zühlke, D., Bernhardt, J., Stülke, J., Hecker, M., 1999. Role of CcpA in regulation of the central pathways of carbon catabolism in *Bacillus subtilis*. *J. Bacteriol.* 181, 6996–7004.
- Tojo, S., Kumamoto, K., Hirooka, K., Fujita, Y., 2010. Heavy involvement of stringent transcription control depending on the adenine or guanine species of the transcription initiation site in glucose and pyruvate metabolism in *Bacillus subtilis*. *J. Bacteriol.* 192, 1573–1585. doi:10.1128/JB.01394-09
- Tojo, S., Satomura, T., Morisaki, K., Deutscher, J., Hirooka, K., Fujita, Y., 2005. Elaborate transcription regulation of the *Bacillus subtilis* *ilv-leu* operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA. *Mol. Microbiol.* 56, 1560–1573. doi:10.1111/j.1365-2958.2005.04635.x
- Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A., Walsh, C.T., 2000. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* 407,

215–8. doi:10.1038/35025116

- Trauger, J.W., Kohli, R.M., Walsh, C.T., 2001. Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry* 40, 7092–8.
- Tsuge, K., Akiyama, T., Shoda, M., 2001. Cloning, sequencing, and characterization of the iturin A operon. *J. Bacteriol.* 183, 6265–73. doi:10.1128/JB.183.21.6265-6273.2001
- Tsuge, K., Ohata, Y., Shoda, M., 2001. Gene yerP, involved in surfactin self-resistance in *Bacillus subtilis*. *Antimicrob. Agents Chemother.* 45, 3566–3573. doi:10.1128/AAC.45.12.3566-3573.2001
- Turgay, K., Hamoen, L.W., Venema, G., Dubnau, D., 1997. Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev.* 11, 119–28.
- Tye, A.J., Siu, F.K.Y., Leung, T.Y.C., Lim, B.L., 2002. Molecular cloning and the biochemical characterization of two novel phytases from *Bacillus subtilis* 168 and *B. licheniformis*. *Appl. Microbiol. Biotechnol.* 59, 190–7. doi:10.1007/s00253-002-1033-5
- Ullrich, C., Kluge, B., Palacz, Z., Vater, J., 1991. Cell-free biosynthesis of surfactin, a cyclic lipopeptide produced by *Bacillus subtilis*. *Biochemistry* 30, 6503–8.
- Vaillancourt, F.H., Yeh, E., Vosburg, D.A., Garneau-Tsodikova, S., Walsh, C.T., 2006. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem. Rev.* 106, 3364–78. doi:10.1021/cr050313i
- van Sinderen, D., Luttinger, A., Kong, L., Dubnau, D., Venema, G., Hamoen, L., 1995. comK encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis*. *Mol. Microbiol.* 15, 455–62.
- Vitiello, A., Ishioka, G., Grey, H.M., Rose, R., Farness, P., LaFond, R., Yuan, L., Chisari, F. V., Furze, J., Bartholomeuz, R., 1995. Development of a lipopeptide-based therapeutic vaccine to treat chronic HBV infection. I. Induction of a primary cytotoxic T lymphocyte response in humans. *J. Clin. Invest.* 95, 341–9. doi:10.1172/JCI117662
- Vlamakis, H., Chai, Y., Beaugard, P., Losick, R., Kolter, R., 2013. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nat. Rev. Microbiol.* 11, 157–168. doi:10.1038/nrmicro2960
- Vollenbroich, D., Ozel, M., Vater, J., Kamp, R.M., Pauli, G., 1997. Mechanism of inactivation of enveloped viruses by the biosurfactant surfactin from *Bacillus subtilis*. *Biologicals* 25, 289–97. doi:10.1006/biol.1997.0099
- Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E., Flugel, R.S., 1997. Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr.*

Opin. Chem. Biol. 1, 309–15.

- Wang, L.F., Wong, S.L., Lee, S.G., Kalyan, N.K., Hung, P.P., Hilliker, S., Doi, R.H., 1988. Expression and secretion of human atrial natriuretic alpha-factor in *Bacillus subtilis* using the subtilisin signal peptide. *Gene* 69, 39–47.
- Weart, R.B., Lee, A.H., Chien, A.-C., Haeusser, D.P., Hill, N.S., Levin, P.A., 2007. A metabolic sensor governing cell size in bacteria. *Cell* 130, 335–47. doi:10.1016/j.cell.2007.05.043
- Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., Holak, T.A., 2000. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* 8, 407–18.
- Wei, Y.-H., Wang, L.-F., Changy, J.-S., Kung, S.-S., 2003. Identification of induced acidification in iron-enriched cultures of *Bacillus subtilis* during biosurfactant fermentation. *J. Biosci. Bioeng.* 96, 174–8.
- Wei, Y.H., Lai, C.C., Chang, J.S., 2007. Using Taguchi experimental design methods to optimize trace element composition for enhanced surfactin production by *Bacillus subtilis* ATCC 21332. *Process Biochem.* 42, 40–45. doi:10.1016/j.procbio.2006.07.025
- Wei, Y.H., Wang, L.F., Chang, J.S., 2004. Optimizing iron supplement strategies for enhanced surfactin production with *Bacillus subtilis*. *Biotechnol. Prog.* 20, 979–983. doi:10.1021/bp030051a
- Wels, M., Kormelink, T.G., Kleerebezem, M., Siezen, R.J., Francke, C., 2008. An in silico analysis of T-box regulated genes and T-box evolution in prokaryotes, with emphasis on prediction of substrate specificity of transporters 16. doi:10.1186/1471-2164-9-330
- Westers, H., Dorenbos, R., Van Dijl, J.M., Kabel, J., Flanagan, T., Devine, K.M., Jude, F., Sérór, S.J., Beekman, A.C., Darmon, E., Eschevins, C., De Jong, A., Bron, S., Kuipers, O.P., Albertini, A.M., Antelmann, H., Hecker, M., Zamboni, N., Sauer, U., Bruand, C., Ehrlich, D.S., Alonso, J.C., Salas, M., Quax, W.J., 2003. Genome Engineering Reveals Large Dispensable Regions in *Bacillus subtilis*. *Mol. Biol. Evol.* 20, 2076–2090. doi:10.1093/molbev/msg219
- Whang, L.-M., Liu, P.-W.G., Ma, C.-C., Cheng, S.-S., 2009. Application of rhamnolipid and surfactin for enhanced diesel biodegradation--effects of pH and ammonium addition. *J. Hazard. Mater.* 164, 1045–50. doi:10.1016/j.jhazmat.2008.09.006
- Willenbacher, J., Mohr, T., Henkel, M., Gebhard, S., Mascher, T., Syltatk, C., Hausmann, R., 2016a. Substitution of the native *srfA* promoter by constitutive *Pveg* in two *B. subtilis* strains and evaluation of the effect on Surfactin production. *J. Biotechnol.* 224, 14–17. doi:10.1016/j.jbiotec.2016.03.002
- Willenbacher, J., Mohr, T., Henkel, M., Gebhard, S., Mascher, T., Syltatk, C., Hausmann, R., 2016b. Substitution of the native *srfA* promoter by constitutive *Pveg* in two *B. Subtilis* strains and evaluation of the effect on Surfactin production. *J. Biotechnol.*

224, 14–17. doi:10.1016/j.jbiotec.2016.03.002

- Willenbacher, J., Yeremchuk, W., Mohr, T., Sylđatk, C., Hausmann, R., 2015. Enhancement of Surfactin yield by improving the medium composition and fermentation process. *AMB Express* 5, 145. doi:10.1186/s13568-015-0145-0
- Wray, L. V., Fisher, S.H., 2011. *Bacillus subtilis* CodY operators contain overlapping CodY binding sites. *J. Bacteriol.* 193, 4841–4848. doi:10.1128/JB.05258-11
- Yan, X., Yu, H.J., Hong, Q., Li, S.P., 2008. Cre/lox system and PCR-based genome engineering in *Bacillus subtilis*. *Appl. Environ. Microbiol.* 74, 5556–5562. doi:10.1128/AEM.01156-08
- Ye, R., Kim, J.H., Kim, B.G., Szarka, S., Sihota, E., Wong, S.L., 1999. High-level secretory production of intact, biologically active staphylokinase from *Bacillus subtilis*. *Biotechnol. Bioeng.* 62, 87–96.
- Yeh, M.S., Wei, Y.H., Chang, J.S., 2006. Bioreactor design for enhanced carrier-assisted surfactin production with *Bacillus subtilis*. *Process Biochem.* 41, 1799–1805. doi:10.1016/j.procbio.2006.03.027
- Yonus, H., Neumann, P., Zimmermann, S., May, J.J., Marahiel, M.A., Stubbs, M.T., 2008. Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J. Biol. Chem.* 283, 32484–91. doi:10.1074/jbc.M800557200
- Youssef, N.H., Duncan, K.E., Michael, J., Mcinerney, M.J., 2005. Importance of 3-Hydroxy Fatty Acid Composition of Lipopeptides for Biosurfactant Activity Importance of 3-Hydroxy Fatty Acid Composition of Lipopeptides for Biosurfactant Activity 71, 7690–7695. doi:10.1128/AEM.71.12.7690
- Zhao, J., Cao, L., Zhang, C., Zhong, L., Lu, J., Lu, Z., 2014. Differential proteomics analysis of *Bacillus amyloliquefaciens* and its genome-shuffled mutant for improving surfactin production. *Int. J. Mol. Sci.* 15, 19847–19869. doi:10.3390/ijms151119847

APPENDIX I

VECTORS AND PLASMIDS

pGEM-T Easy

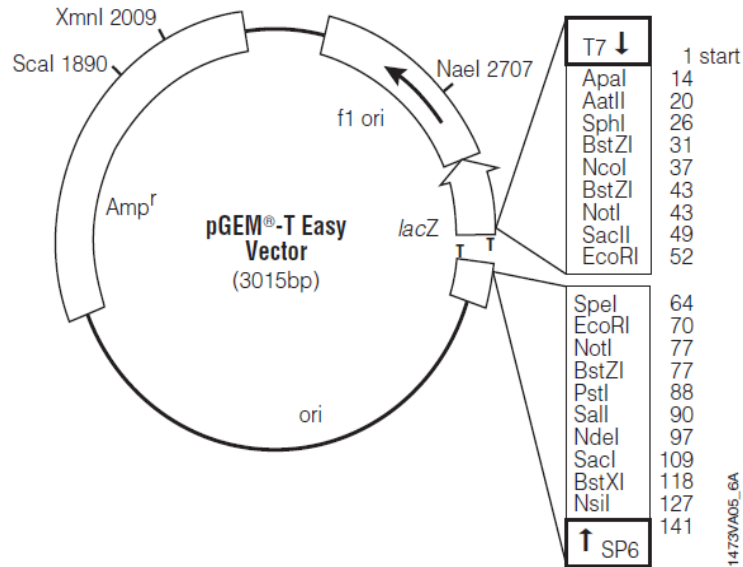


FIGURE pGEM-T Easy vector with different restriction sites supplied by Promega

pDG1661

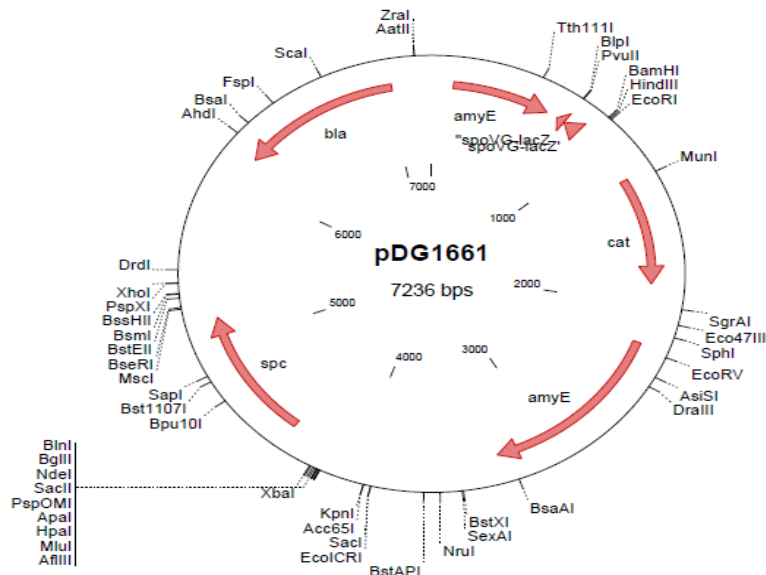


FIGURE Plasmid pDG1661 (*amyE-spoVG-lacZ-amyE*) used for the construction of pBG400

pBC16

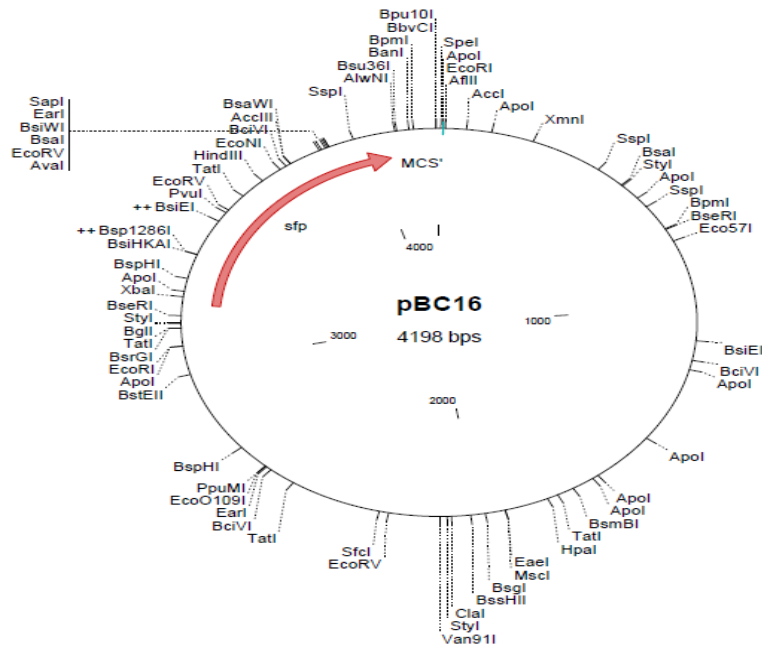


FIGURE Plasmid pBC16 (*sfp*) used for the construction of pBG400

pUC57

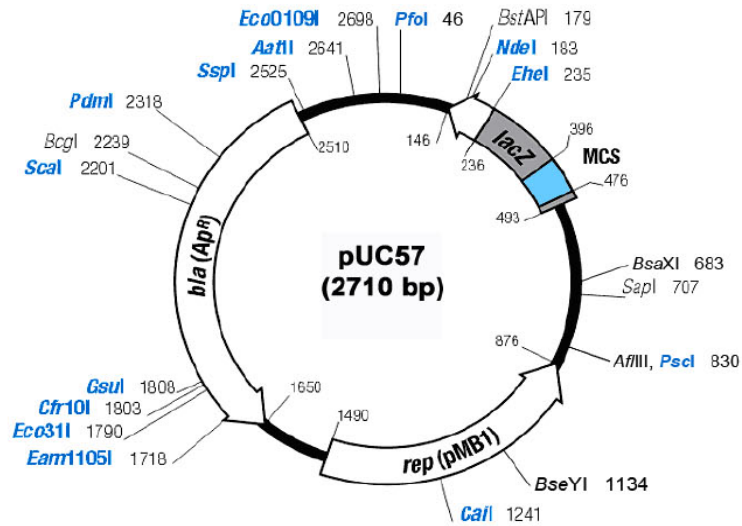


FIGURE The synthesized sequence ($P_{repU-ermC}$) was inserted in the MCS of pUC57; the vector was shown supplied by GenScript

pBG124

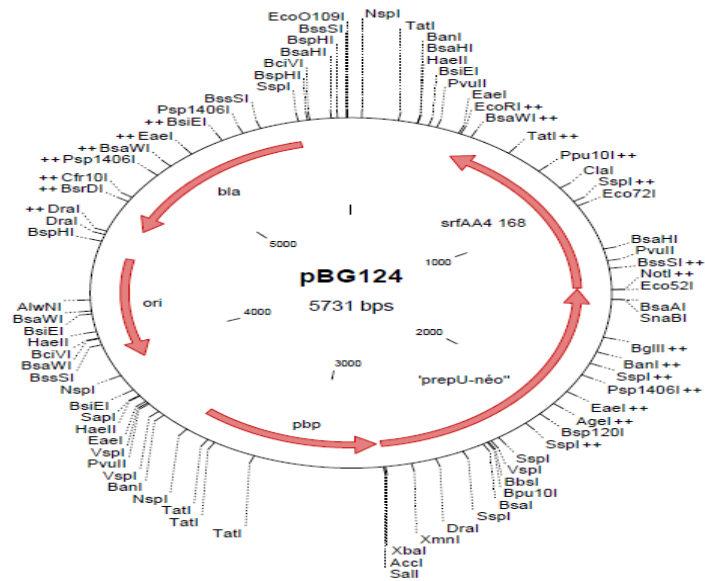


FIGURE Plasmid pBG124 containing the $P_{repU-neo}$ sequence along with *Hae*III restriction site and *srfAA* sequence used for the construction of pBG133

pBG127

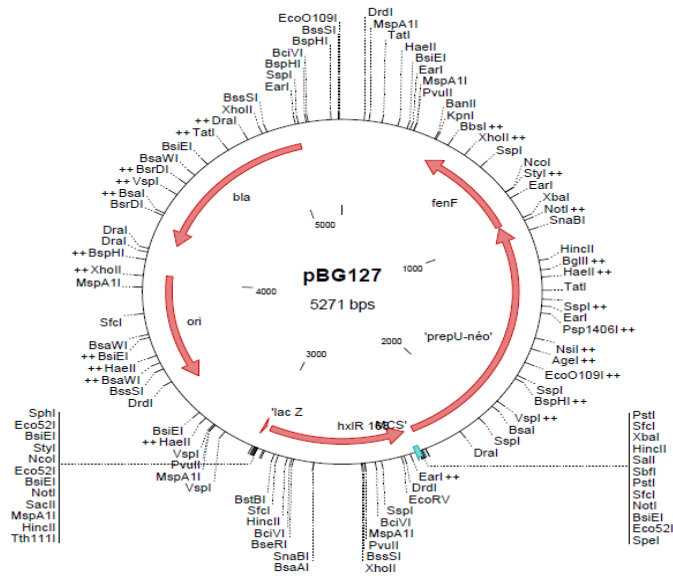


FIGURE Plasmid pBG127 containing *hxlR* sequence used for the construction of pBG133

Synthesized sequence (P_{repU} - *ermC* - P_{srfAA})

AAGCGCCGGCGGACGTC**GTCGAC**AAGTCTCGAACCCGACGTCGAATTCGAGCTCGGTAATCAA
GCGGCCGAAGGGGATCCTCTAGAGCTTGGGCTGCAGGTCGAAAAAGAAATTAGATAAATC
TCTCATATCTTTTATTCATAATCGCATCCGATTGCAGTATAAATTTAACGATCACTCATCA
TGTTTCATATTTATCAGAGCTCGTGCTATAATTATACTAATTTTATAAGGAGGAAAAAATAT
GGGCATTTTTAGTATTTTTGTAATCAGCACAGTTCATTATCAACCAAACAAAAAATAAGTG
GTTATAATGAATCGTTAATAAGCAAAATTCATATAACCAAATTAAGAGGGTTATAATGAA
CGAGAAAAATATAAAACACAGTCAAACTTTATTACTTCAAAACATAATATAGATAAAATA
ATGACAAATATAAGATTAATGAACATGATAATATCTTTGAAATCGGCTCAGGAAAAGGC
CATTTTACCCTTGAATTAGTAAAGAGGTGTAATTTTCGTAACCTGCCATTGAAAATAGACCATA
AATTATGCAAACTACAGAAAAATAAATTGTTGATCACGATAATTTCCAAGTTTTAAACAA
GGATATATTGCAGTTTTAAATTTCTAAAAACCAATCCTATAAAATATATGGTAATATACCT
TATAACATAAGTACGGATATAATACGCAAAATGTTTTTTGATAGTATAGCTAATGAGATTT
ATTTAATCGTGGAATACGGGTTTGCTAAAAGATTATTAATACAAAACGCTCATTGGCATT
ACTTTTAAATGGCAGAAGTTGATATTTCTATATTAAGTATGGTTCCAAGAGAATATTTTCAT
CCTAAACCTAAAGTGAATAGCTCACTTATCAGATTAAGTAGAAAAAATCAAGAATATCAC
ACAAAGATAAAACAAAAGTATAATTATTTTCGTTATGAAATGGGTTAACAAAGAATACAAGAA
AATATTTACAAAAAATCAATTTAACAATTCCTTAAAAACATGCAGGAATTGACGATTTAAAC
AATATTAGCTTTGAACAATCTTATCTCTTTTCAATAGCTATAAATATTTAATAAGTAAGT
TAAGGGATGCATAAACTGCATCCCTTAACTTGTTTTTCGTGTGCCTATTTTTTGTGAATCG
ATTATGTCTTTTGCAGCTCTCGAATACGAATCCTTCTGCTCAATAATTAT**GGTACC**ACGAGG
TGAAATCAATTGAGTATCAATATAAACTTTATATGAACATAATCAACGAGGTGAAATCGAGAT
CAGGGAATGAGTTTATAAAAAATAAAAAAGCACCTGAAAAGGTGTCTTTTTTTGATGGTTTT
GAACTTGTTCTTTCTTATCTTGATACATATAGAAATAACGTCATTTTTATTTTAGTTGCTGA
AAGGTGCGTTGAAGTGTGGTATGTATGTGTTTTAAAGTATTGAAAACCTTAAAAATTGGT
TGCACAGAAAAACCCATCTGTAAAGTTATAAGTGACTAAACAAATAACTAAATAGATGG
GGGTTTCTTTAATATTATGTGTCCTAATAGTAGCATTATTCAGATGAAAAATCAAGGGT
TTTAGTGGACAAGACAAAAAGTGAAAAAGTGAGACCATGGCTAGCGAATTCAGTAGTGATT
TTTAATAGCGGCCATCTGTTTTTTGATTGGAAGCACTGCTTTTTAAGTGTAGTACTTTGGG
CTATTTTCGGCTGTTAGTTCATAAGAATTAAGGCTGATATGGATAAGAAAGAGAAAATGC
GTTGCACATGTTCACTGCTTATAAAGATTAGGGGAGGTATGACAATATGGAAATAACTTTT
TACCCTTTAACGGATGCACAAAAACGAATTTGGTACACAGAAAAATTTATCCTCACACGA
GCATTTCAAATCTTGCAGGGGATTGGTAAGCTGGTTTCAGCTGATGCGATTGATTATGTGC
TTGTTGAGCAGGCGATTCAAGAGTTTATTCGCAGAAATGACGCCATGCGCCTTCGGTTGC
GGCTAGATGAAAACGGGGAGCCTGTTCAATATATTAGCGAGTATCGGCCTGTTGATATAA
AACATACTGACACTACTGAAGATCCGAATGCGATAGAGTTTATTTACAATGGAGCCGGG
AGGAAACGAAGAAACCTTTGCCGCTATACGATTGTGATTTGTTCCGTTTTTCCTTGTTTAC
CATAAAGGAAAATGAAGTGTGGTTTTACGCAATGTTTAT**CACCTG**ATTTCTGATGGTATG
TCCATGAAT

GTCGAC : RE SITE *AccI* (*XmiI*)

GAG....ATG : P_{repU}

GGTACC : RE SITE *KpnI*

GAA...GTC : *erm*^R

TTT....ATC : *HaeIII-srfAA*

CACCTG : RE SITE *Eco72 I* (*PmlI*)

APPENDIX II

MEDIA COMPOSITIONS

LB MEDIUM

COMPONENTS	CONCENTRATION
Tryptone	10 g/L
Yeast extract	5 g/L
NaCl	10 g/L

COMPETENCE MEDIUM

A. Competence medium (completed MC) – 2ml

COMPONENTS	VOLUME	TREATMENT
MQ water	1.8 ml	-
10X MC (B)	200 µl	filter sterilized
1M MgSO ₄	6.7 µl	autoclaved
1% Tryptophan (Optional based on BS168 or BSB1 strain)	10 µl	filter sterilized (stored in aluminum foil)

- B. MC 10X** – The total volume of the mixture was 100 ml. Add 50 ml water and then add the components and finally adjust the volume to 100ml. Adjust pH to 7. Filter sterilize at store at -20°C.

COMPONENTS	CONCENTRATION / VOLUME
K ₂ HPO ₄	140.4 g/l
KH ₂ PO ₄	52.4 g/l
Glucose	200 g/l
300mM Tri-Na citrate (wrap in aluminium foil and store at -20°C)	10 ml
Ferric NH ₄ citrate (wrap in aluminium foil and store at -20°C)	1 ml
Caesin hydrolysate	10 g/l
Potassium Glutamate	20 g/l

LANDY MEDIUM

- A. Classical Landy medium**– Required volume of water was added prior to addition of nutrients

COMPONENTS	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME (1 L)
Solution A: Yeast Extract	20 g/l	1 g/l	50 ml
MgSO ₄	10 g/l	0.5 g/l	
Solution B: K ₂ HPO ₄	20 g/l	1 g/l	50 ml
KCl	10 g/l	0.5 g/l	
Solution C: CuSO ₄	0.032 g/l	0.0016 g/l	50 ml
MnSO ₄	0.024 g/l	0.0024 g/l	
FeSO ₄	0.008 g/l	0.0008 g/l	
Glucose	400 g/l	20 g/l	50 ml
Glutamic acid	100 g/l	5 g/l	50 ml
MOPS	420 g/l	21 g/l	50 ml
L-Tryptophane (Optional based on BS168 or BSB1 strain)	1.6 g/l	0.016	10 ml

B. Modified Landy medium with MOPS buffer- Required volume of water was added prior to addition of nutrients

COMPONENTS	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME (1 L)
Solution A: Yeast Extract	20 g/l	1 g/l	50 ml
MgSO ₄	10 g/l	0.5 g/l	
Solution B: K ₂ HPO ₄	20 g/l	1 g/l	50 ml
KCl	10 g/l	0.5 g/l	
Solution C: CuSO ₄	0.032 g/l	0.0016 g/l	50 ml
MnSO ₄	0.024 g/l	0.0024 g/l	
FeSO ₄	0.008 g/l	0.0008 g/l	
Glucose	400 g/l	20 g/l	50 ml
(NH ₄) ₂ SO ₄	72 g/l	3.6 g/l	50 ml
MOPS	420 g/l	21 g/l	50 ml
L-Tryptophane (Optional based on BS168 or BSB1 strain)	1.6 g/l	0.016	10 ml

C. Modified Landy medium with Potassium phosphate buffer – Required volume of water was added prior to addition of nutrients

COMPONENTS	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME (1 L)
Yeast Extract	20 g/l	1 g/l	50 ml
MgSO ₄	246.5 g/l	0.25 g/l	1 ml
KCl	10 g/l	0.5 g/l	50 ml
Glucose	400 g/l	20 g/l	50 ml
K – BUFFER: K ₂ HPO ₄	122 g/l	12.2 g/l	100 ml
KH ₂ PO ₄	41 g/l	4.1 g/l	
NaCl	8.8 g/l	0.88 g/l	
(NH ₄) ₂ SO ₄	72 g/l	3.6 g/l	50 ml
Solution C: CuSO ₄	0.032 g/l	0.0016 g/l	50 ml
MnSO ₄	0.024 g/l	0.0024 g/l	
FeSO ₄	0.008 g/l	0.0008 g/l	

TSS SUPPLEMENTED WITH 16 AMINO ACIDS

A. Entire mixture - Required volume of water was added prior to addition of nutrients

COMPONENTS	FINAL CONCENTRATION
Tris	6.05 g/l
FeCl ₃	0.04 g/l
K ₂ HPO ₄	0.44 g/l
MgSO ₄	0.2 g/l
NH ₄ Cl	2 g/l
Glucose	5 g/l
16 Amino acids (B)	4.5 g/l
MOPS	21 g/l

B. 16 Amino acids

COMPONENTS	FINAL CONCENTRATION
L-cysteine	0.04 g/l
L-arginine	0.4 g/l
L-isoleucine	0.2 g/l
L-leucine	0.2 g/l
L-valine	0.2 g/l
L-glutamic acid	0.8 g/l
L-aspartic acid	0.665 g/l
L-lysine	0.1 g/l
L-phenylalanine	0.1 g/l
L-proline	0.1 g/l
L-threonine	0.1 g/l
L-alanine	0.445 g/l
Glycine	0.375 g/l
L-serine	0.525 g/l
L-tryptophan	0.15 g/l
L-methionine	0.16 g/l

AMBER MEDIUM - Required volume of water was added prior to addition of nutrients

COMPONENTS	STOCK CONCENTRATION	COMMENTS	FINAL CONCENTRATION	VOLUME (1 L)
K ₂ HPO ₄	122 g/l	Buffer solution pH 7.0-7.1	12.2 g/l	100 ml
KH ₂ PO ₄	41 g/l		4.1 g/l	
NaCl	8.8 g/l		0.88 g/l	
(NH ₄) ₂ SO ₄	132 g/l	Autoclave	1.32 g/l	10 ml
ZnCl ₂	0.27 g/l	Trace Elements Dissolve in water one by one in this order Filtrate 4°C	0.00027 g/l	1 ml
MnSO ₄ ·H ₂ O	0.34 g/l		0.00034 g/l	
CuCl ₂ ·2H ₂ O	0.34 g/l		0.00034 g/l	
CoCl ₂ ·6H ₂ O	0.47 g/l		0.00047 g/l	
Na ₂ MoO ₄ ·2H ₂ O	0.48 g/l		0.00048 g/l	
D-glucose	400 g/l	Filtrate 4°C	4 g/l	10 ml
L-malate	500 g/l	Filtrate 4°C	3 g/l	6 ml
Tryptophan	5 g/l	Filtrate 4°C (Optional based on BS168 or BSB1 strain)	0.05 g/l	10 ml
K-glutamate	203 g/l	Filtrate 4°C	2.03 g/l	10 ml
MgSO ₄ ·7H ₂ O	247 g/l	Autoclave	2.47 g/l	1 ml
CaCl ₂ ·2H ₂ O	14.7 g/l	Filtrate 4°C	0.0147 g/l	1 ml
Fe-NH ₄ -citrate	2.2 g/l	Filtrate 4°C (keep in dark)	0.00264 g/l	1.2 ml

APPENDIX III

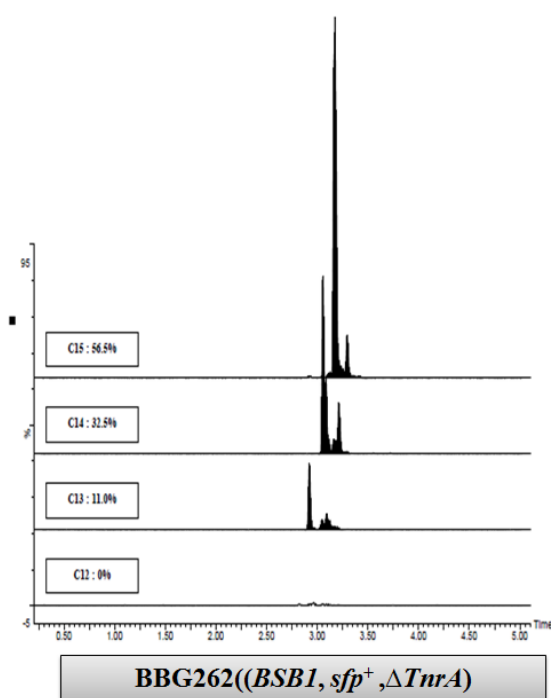
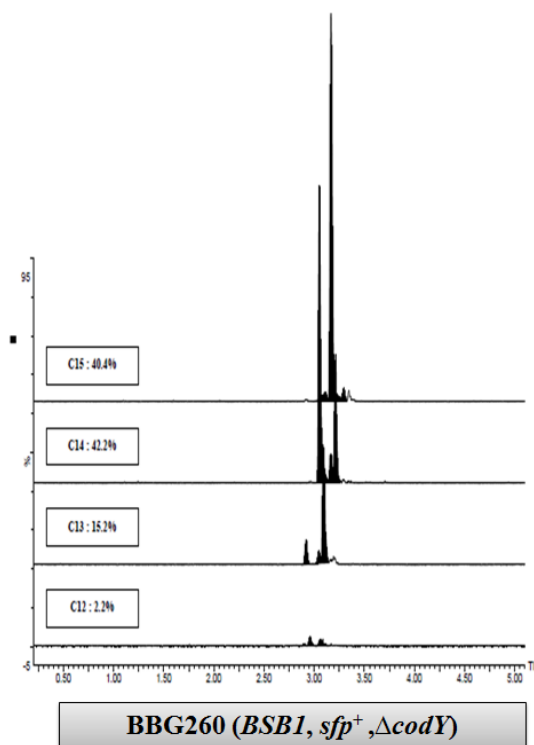
Molecules of branched chain amino acid metabolic network (Figure 34)

ROLE	ABBREVIATIONS	CHEMICAL SPECIES	
Metabolites	Ile	Isoleucine	
	Leu	Leucine	
	Val	Valine	
	Akb	L-2-amino-acetoacetate	
	Glu	Glutamate	
	OxoGlu	Oxoglutarate	
	GTP	Guanosine triphosphate	
	Ket _a	2-keto-3-methylvalerate	
	Acyl-CoA	Acyl Coenzyme A	
	Ket _b	2-keto-isovalerate	
	Ket _c	2-keto-isocaproate	
	Pyr	Pyruvate	
	Thr	Threonine	
	Proteins	Bcd	Branched chain amino-acid dehydrogenase
		BkL	2-oxoisovalerate dehydrogenase
BkdR		Transcriptional activator of BkL	
CcpA		Carbon catabolite control protein A	
CodY		Transcriptional pleiotropic regulator	
IlvA		Threonine deshydratase	
IlvBH		Acetolactate synthase	
IlvC		Ketol-acid reductoisomerase	
IlvD		Dihydroxy-acid dehydratase	
LeuA		2-isopropylmalate synthase	
LeuBCD		3-isopropylmalate dehydratase	
TnrA		Nitrogen pleiotropic transcriptional regulator	
YbgE		Branched chain amino-acid aminotransferase	
YwaA		Branched chain amino-acid aminotransferase	
Actors		P _{Ilv-Leu}	Activity of promoter of IlvBH IlvC LeuA LeuBCD operon
	BS _{CodY}	Activity of CodY binding to promoter P _{Ilv-Leu}	
	BS _{TnrA}	Activity of TnrA binding to promoter P _{Ilv-Leu}	
	BS _{CcpA}	Activity of CcpA binding to promoter P _{Ilv-Leu} without BS _{TnrA} loop	
	OP _{BkL-Bcd}	Activity of promoter of BkL Bcd operon	
	YwaA+YbgE	Joint activity of YbgE and YwaA	
Tbox	Activity of leucine attenuation		

Reactions of branched chain acid network (Figure 34)

Name	Function
r1	bind CodY to $P_{Ilv-Leu}$ for inhibition
r2	activate $P_{Ilv-Leu}$ promoter
r3	bind BkdR to BkL Bcd promoter
r4	constitutive expression of BkL Bcd operon
r5	express YbgE
r6	express YwaA
r7	bind TnrA to $P_{Ilv-Leu}$ promoter for inhibition
r9	bind CcpA to $P_{Ilv-Leu}$ promoter without BS_{TnrA} loop
r10	expression of Bcd, activated by $OP_{BkL-Bcd}$
r11	expression of BkL, activated by $OP_{BkL-Bcd}$
r12	activate BkdR by Ile
r13	activate BkdR by Val
r14	expression of CcpA
r15	express and accelerate CodY (to be explained acceleration)
r16	expression of TnrA
r17	Gtp degradation
r18	BkdR degradation
r19	YwaA+YbgE degradation
r20	disabling of IlvA by Ile
r21	disabling of IlvA by Val
r22	expression of IlvA inhibited by binding of CodY
r23	expression of IlvBH
r24	expression of IlvC
r25	expression of IlvD, inhibited by binding of CodY to promotor
r26	metabolic transformation of Ket_a to Ile activated by YwaA+YbgE, after an amino addition taken from Glu, which becomes OxoGlu.
r27	metabolic transformation of Akb and Pyr into Ket_a
r28	prepare outflow of Ket_a activated by BkL
r29	degradation of Ile into Ket_a activated by YwaA and Bcd, with an amino transfer from OxoGlu which becomes Glu.
r30	metabolic transformation from Pyr to Ket_b , activated by IlvD
r31	prepare outflow of Ket_b activated by BkL
r32	degradation of Val into Ket_b activated by YwaA and Bcd with an amino transfer from OxoGlu which becomes Glu.
r33	metabolic transformation of Ket_b into Ket_c , activated by LeuBCD and LeuA
r34	prepare outflow of Ket_c activated by BkL
r35	degradation of Leu into Ket_c activated by YwaA and Bcd with an amino transfer from OxoGlu which becomes Glu.
r36	metabolic transformation of Ket_b to Val activated by YwaA+YbgE, after an amino addition from Glu which becomes OxoGlu
r37	expression of LeuA
r38	expression of LeuBCD
r39	deactivation of LeuA by Leu
r40	Leucine attenuation
r41	metabolic transformation of Thr into Akb using IlvA
r45	metabolic transformation of Ket_c to Leu activated by YwaA+YbgE after an amino addition from Glu, which becomes OxoGlu
r46	expression of YbgE, inhibited by binding of CodY to promoter
r47	expression of YwaA

LC-MS-MS analysis



APPENDIX IV

Research Article

Modeling leucine's metabolic pathway and knockout prediction improving the production of surfactin, a biosurfactant from *Bacillus subtilis*

François Coutte^{1,2,*}, Joachim Niehren^{3,4,*}, Debarun Dhali^{1,2}, Mathias John^{2,3}, Cristian Versari^{2,3} and Philippe Jacques^{1,2}

¹ ProBioGEM team, Research Institute for Food and Biotechnology - Charles Viollette (EA7394), University of Lille, Villeneuve d'Ascq, France

² University of Lille, Villeneuve d'Ascq, France

³ BioComputing team, CRISTAL Lab (CNRS UMR9189), University of Lille, Villeneuve d'Ascq, France

⁴ Inria Lille, Villeneuve d'Ascq, France

A *Bacillus subtilis* mutant strain overexpressing surfactin biosynthetic genes was previously constructed. In order to further increase the production of this biosurfactant, our hypothesis is that the surfactin precursors, especially leucine, must be overproduced. We present a three step approach for leucine overproduction directed by methods from computational biology. Firstly, we develop a new algorithm for gene knockout prediction based on abstract interpretation, which applies to a recent modeling language for reaction networks with partial kinetic information. Secondly, we model the leucine metabolic pathway as a reaction network in this language, and apply the knockout prediction algorithm with the target of leucine overproduction. Out of the 21 reactions corresponding to potential gene knockouts, the prediction algorithm selects 12 reactions. Six knockouts were introduced in *B. subtilis* 168 derivatives strains to verify their effects on surfactin production. For all generated mutants, the specific surfactin production is increased from 1.6- to 20.9-fold during the exponential growth phase, depending on the medium composition. These results show the effectiveness of the knockout prediction approach based on formal models for metabolic reaction networks with partial kinetic information, and confirms our hypothesis that precursors supply is one of the main parameters to optimize surfactin overproduction.

Received	15 FEB 2015
Revised	26 MAY 2015
Accepted	09 JUL 2015
Accepted article online	14 JUL 2015

Supporting information
available online



Keywords: Abstract interpretation · *Bacillus subtilis* · Knockout prediction · Modeling language · Surfactin

1 Introduction

In this paper we develop and implement genetic engineering methods with the objective to overproduce surfactin by *Bacillus subtilis*. Surfactin is one of the most powerful biosurfactant known and it displays several biological activities of interest (antiviral, antimycoplasmic, elicitor, etc) [1]. This promising molecule is assembled by a nonribosomal mechanism involving multifunctional proteins called nonribosomal peptide synthetases (NRPS). The substrates of these enzymes are amino acids or fatty acids residues present in the cytoplasm of the cell. The

* These authors contributed equally to this work.

Corresponding author: Dr. François Coutte, Research Institute for Food and Biotechnology – Charles Viollette, Polytech-Lille, Université de Lille, Sciences et Technologies, 59655 Villeneuve d'Ascq, France.
E-mail: francois.coutte@polytech-lille.fr

Abbreviations: Acyl-CoA, acyl coenzyme A; Akb, L-2-amino-acetoacetate; BCAA, branched chain amino acid; Glu, glutamate; Ctp, guanosine triphosphate; Ile, isoleucine; Ket₃, 2-keto-3-methylvalerate; Ket₂, 2-keto-isovalerate; Ket₁, 2-keto-isocaproate; Leu, leucine; NRPS, nonribosomal peptide synthetase; OxoGlu, oxoglutarate; P_{ilv-Leu}, *ilv-leu* operon promoter; Pyr, pyruvate; Thr, threonine; Val, valine; XML, eXtensible Markup Language; XSLT, eXtensible Stylesheet Language Transformations

2015 The Authors. Biotechnology Journal published by Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1216

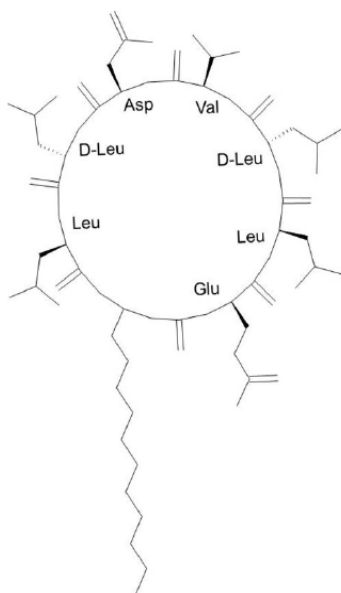


Figure 1. Schematic representation of surfactin [2].

surfactin is composed of a ring of seven amino acid residues connected to a β -hydroxylated fatty acid chain of different length and isomery [1, 2]. The peptide moiety contains four leucines (Fig. 1). Genetic engineering of *B. subtilis* has already been made in order to increase the lipopeptide production. In previous work [3], the overproduction of surfactin was obtained by replacing the native promoter of the surfactin operon (*urfA*) by a constitutive one and disrupted the plipastatin operon (*ppsA*) to save the precursor availability. The same approach was recently developed for the mycosubtilin production [4]. Sun et al. [5] replaced the *urfA* promoter by P_{spac} . Liu et al. [6] explore feeding strategies of amino acid to modify the peptide moiety of the lipopeptide. However, there is no work directly related to the orientation of the metabolism to increase the supply precursor without modifying the extracellular medium. The hypothesis of the present work is that any further increase of surfactin production can only be achieved by increasing the production of these precursors, and in particular leucine.

The branched chain amino acids, i.e. leucine, valine, and isoleucine are produced by a metabolic pathway with complex regulatory mechanisms. The question is how to adapt this pathway of *B. subtilis* by gene knockouts, in order to overproduce leucine. Our hope is that one can predict such gene knockouts from models of the pathway by methods from computational biology, in order to narrow down the testing duties in wet lab. Model-based

knockout prediction for metabolic pathways, however, requires to overcome a number of difficulties.

Flux balance analysis is the most prominent approach for model-based prediction [7–9]. The idea is to infer flux balance equations from the models [10], and to analyze the space of their solutions. Flux balance equations are linear equations with variables for fluxes, i.e. real numbers that design the speed of a reaction in a steady state. The problem is that the number of variables exceeds the number of equations in the system by far, so that the spaces of solutions are huge. Therefore, particular solutions must be filtered, typically by imposing optimization criteria in various manners. One can try to optimize the biomass [11–13], that is the production of metabolites that are needed for an optimal growth of the cells. Or else, one may apply metabolic flux analysis [14] in order to measure experimentally the sizes of a sufficient number of metabolic flows. However, the specific growth conditions used in the experiments strictly limit the applicability of the results. The advantage of flux balance analysis is that it can be applied to models of reaction networks without any kinetic information. But this is also problematic, since flux balance equations cannot express any kinetic information. Therefore, they cannot distinguish the various kinds of modifiers such as enzymes and inhibitors that are positive from negative regulatory effects, which are available in more recent models of metabolic networks [15]. On the other hand, these models lack any clear formal semantics which makes them useless for formal reasoning. Constraint-based approaches such as [12, 13] derive Boolean constraints from reaction networks, which mainly state the dependencies of where a reaction is on or off from the presence of its reactants or modifiers. Such Boolean constraints can be seen as an abstraction of the steady state equations used by flux balance analysis. But the under-constrainedness problem remains as with flux balance analysis, so that one needs to impose further restrictions on the solution space. This can be done for instance by adding negative constraints stating that the reaction is off if one of its inhibitors is present. Note that such negative constraints are often too strong, since an inhibitor often slows down a reaction in a steady state, but does not shut it down totally at least not in average.

Alternatively, pathway analysis techniques may help to discover genetically independent pathways [16, 17]. Among these, elementary mode analysis is also based on flux balance equations. The idea is to change the basis of the flux balance equations, so that they constrain the speeds of pathways, i.e. of paths from an input to an output in a reaction graph, rather than the speeds of the reactions themselves. When adding Boolean constraints in addition, one can express information about inhibitors of reactions, too [18]. This method is known to be problematic for networks with many pathways, that applies to metabolic networks with many paths from inputs to outputs, such as in our case.

A more systematic analysis of reaction networks with partial kinetic information was proposed in [19]. Steady state equations are inferred from the reaction networks, while the usual flux balance equations contain constrained variables for the partially known kinetic functions. Abstract interpretation is then applied in order to abstract away the variables on the kinetic knowledge. It produces difference constraints, that can state that the speed of a reaction will increase if the concentration of one of its inhibitors decreases, and conversely. This is contrast to the previous treatments of inhibitors by Boolean constraints, where one can only reason about reactions being on or off. Difference constraints are particular finite domain constraints. This makes it possible to use finite domain constraint solving to predict gene knockouts that will speed up a target reaction. The approach of [19], however, is too restrictive since it is limited to networks of reactions with mass action kinetics, for which the rate constants are unknown. In order to overcome this restriction, Niehren et al. [20] proposed a modeling language for reaction networks with partial kinetic information, which supports unknown kinetic functions up to a similarity constraint that captures activating or inhibiting effects. Reaction networks modeled in this language can still be analyzed by abstract interpretation while generalizing on [19]. And the finite domain constraints inferred from the models can be used to predict flux changes based on finite domain constraint solving. However, no knockout prediction algorithm along these lines has been proposed yet.

As a first contribution in this article, we develop a knockout prediction algorithm for reaction networks in this modeling language. Basically, the algorithm is obtained by the combination of the ideas of the two previous papers [19, 20]. For the sake of self-containedness, we elaborate these ideas from scratch, and illustrate the algorithm obtained at the example reaction network modeling the regulation of the *ilv-leu* operon promoter ($P_{ilv-leu}$). As we argue, its correctness can be reduced to the previous results. Our algorithm is based on finite-domain constraint solving to enumerate the solutions of a difference constraint obtained by abstract interpretation. The maximal number of changes in a solution must be fixed by the user. From the set of solutions obtained, the algorithm then filters those that correspond to single knockouts in a straightforward manner. The result is a list of reactions that are predicted for single knockout. All this can be done fully automatically, when given the reaction network, the knockout candidates, the overproduction target, and the maximal number of changes in a solution.

The second contribution is an implementation of all these algorithms. This is obtained by extending the finite domain constraint solver for difference constraints from [19], which is written in SCALA, as well as the tools for the

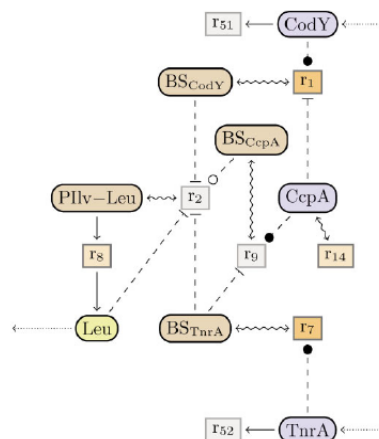


Figure 2. Reaction network of $P_{ilv-leu}$ regulation. The concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm.

modeling language from [20], which are written in XSLT. More precisely, these tools allow the validation of reaction networks in XML, convert them into a nice graphical output (e.g. Fig. 2 and 3), compute the steady state equations, and infer the difference constraints by abstract interpretation. Finally, the knockout predictions are obtained by filtering the solutions of the difference constraints corresponding to single knockouts by yet another XSLT script.

Our third contribution is a model of the quite complex leucine metabolic pathway [21–25] in the modeling language from [20]. This is obtained by rewriting the model from SubtiWiki [15, 26] into this formal modeling language, while adding few missing reactions. We then applied the new prediction algorithm to this model. It predicts 14 reaction knockouts from the 21 given candidates, but only for 12 of these reactions, the effect of a knockout is plausible with respect to the model. It might be possible to rule out the two implausible predictions based on the model too, but for now we do not know how to do this automatically.

Our fourth contribution is a successful experimental verification in vivo of the six predicted knockouts with the most direct effects by genetic engineering of *B. subtilis*. It turns out that for all generated mutants the specific surfactin production is increased by a factor ranging from 1.6 to 20.9 during the exponential growth phase, depending on the gene knockout and on the medium composition. These results show the effectiveness of the knockout prediction and confirm our hypothesis that precursors supply is one of the main parameter to optimize surfactin overproduction.

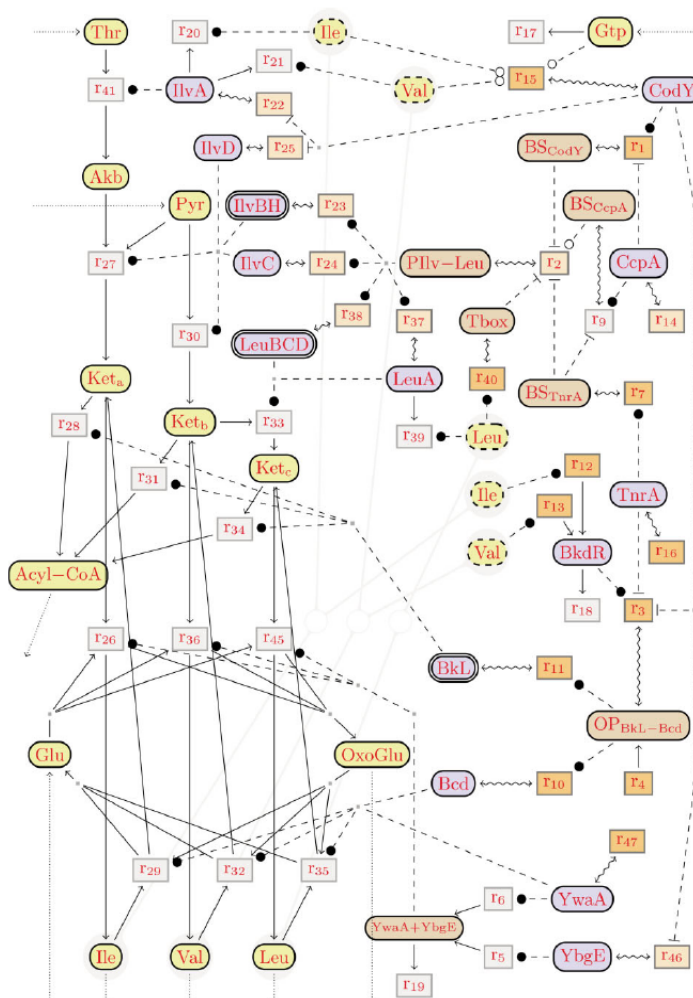


Figure 3. Branched chain amino acid reaction network. The concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm.

2 Materials and methods

2.1 In-vivo experiments and genetic engineering of *B. subtilis*

2.1.1 Metabolic effect of leucine feeding

In order to test the working hypothesis that an increase in the intracellular pool of leucine leads to an increase of the surfactin production, the strain *B. subtilis* BBG111 [3] was cultivated in presence of leucine. Two different cultures were carried-out in Erlenmeyer flask at 37°C with 160 rpm of agitation. The first medium used was Landy [27] with 3.6 g/L $(\text{NH}_4)_2\text{SO}_4$ (instead of glutamate) and the second

one Landy with 2.3 g/L $(\text{NH}_4)_2\text{SO}_4$ and 2.5 g/L of leucine (instead of glutamate). Medium were supplemented with tryptophan (16 mg/L) and buffered at pH 7.0 with MOPS 0.1 M. Samples were withdrawn after 6 h of culture. Biomass and optical density measurement were carried out as described previously [3]. Surfactin production was measured by RP-HPLC in the centrifugated and filtered samples, according to Coutte et al. [3] and using pure surfactin as the standard (Lipofabrik, Villeneuve d'Ascq, France). Cultures were performed in triplicate and results are expressed as mean value with standard deviation. The results given in section 3.5 confirmed the working hypothesis.

2.1.2 Genetic engineering of *B. subtilis* 168 derivative strains

We introduced gene knockout in the leucine metabolic pathway using the pop-in pop-out technique [28]. Construction of all deletion mutant strains were carried out as described in Supporting information, Table S1. A modified *B. subtilis* BSB1 strain [29] was constructed using the genomic DNA of TF8A strain by introducing a neomycin resistance gene under the control of the Lambda promoter (P_{λ} -neo). All deletions were introduced in the master strain by homologous replacement of the targeted chromosomal region by a DNA fragment called "cassette *upp-phleo-cl*", carrying the phleomycin-resistance gene for positive selection of cassette integration and $P_{sak-Acl}$ genes for counterselection. After deletion of a particular fragment the chromosomal arrangement was confirmed by replica plate method as well as through PCR analysis and gel electrophoresis on 0.8% agarose gel.

Then, two sets of primer pairs were designed for upstream and downstream region of the gene to be deleted (using the sequence from SubtiList World-Wide Web Server and supplied by Eurofins Genomics, Ebersberg, Germany). These two sets of primers were named (forward₁,reverse₁) for the upstream gene and (forward₂,reverse₂) for the downstream gene respect to the gene of interest. During the design of the primers sets, direct repeat sequence are either added to the primer sequence reverse₁ or forward₂ as well as complement sequence of 5' and 3' region of the cassette *upp-phleo-cl* are also added to the primer sequence of reverse₁ and forward₂ (Supporting information, Table S2). Polymerase chain reaction was carried out using specific primer pair (forward₁,reverse₁) and (forward₂,reverse₂) (0.5 μ M final each) and 200 ng of master strain chromosomal DNA under standard condition to produce DNA fragments which were at least 1.0 kb long. The primers used for amplification are listed in Table S2. The PCR was carried out under following condition: 5 min at 94°C; (30 s at 94°C, 30 s at 55°C, 2 min at 72°C) for 25 cycles; 5 min at 72°C. The PCR products were purified using gel extraction kit (Qiagen, Hilden, Germany). The DNA fragments generated using (forward₁,reverse₁) and (forward₂,reverse₂) was mixed equally with higher proportion of *upp-cl-phleo* cassette and joining PCR reaction was conducted under the following conditions: 5 min at 94°C; (15 s at 94°C, 15 s at 55°C, 12 min at 65°C) for 12 cycles; (15 s at 94°C, 15 s at 55°C, 12 min at 65°C) for 25 cycles; 10 min at 72°C in the presence of forward₁ and reverse₂ (0.1 μ M final). The resultant product was confirmed through electrophoresis on 0.8% agarose gel and further used to transform competent cells of the master strain.

Positive selection of deletion mutants was carried out through phleomycin resistance on LB medium plates, which were incubated at 37°C up to 24 h. Colonies, with proper cell morphology were streaked twice on the same medium along with on neomycin-LB medium plates in

order to purify the colony. Each clone was checked for the absence of the deleted chromosomal region through PCR using primer pair forward₁ and reverse₂ respectively. Strains constituting the expected deleted chromosomal structure were preserved in glycerol stock at -80°C.

All deletions obtained in modified *B. subtilis* BSB1 strain were introduced in *B. subtilis* BBG111 [3] through transformation of genomic DNA of deletion strain into *B. subtilis* BBG111 through natural competence transformation [30]. The deletion was confirmed through replica plate method using phleomycin, neomycin and phleomycin/neomycin antibiotic plates (Phleomycin was supplied by Euromedex, Souffelweyersheim, France, and Neomycin by Sigma-Aldrich, St Louis, USA). The constructions of the mutant strains are provided in Supporting information, Table S1. This procedure was followed for the knockout of the genes *bcd*, *codY*, *bkL* (whenever we talk about a knockout for the *bkL* operon, we mean more precisely the *lpdV* gene of this operon) and *tnrA*, giving rise to mutants strains named *B. subtilis* BBG253, BBG254, BBG255 and BBG256.

Two remaining strains were obtained from the strains constructed in the Sonenshein Lab [24]. The first modified one was SRB94, in which site directed mutagenesis has been carried out to generate A10C and C7G mutation in the region II of CodY binding site at *ilv-leu* promoter. The second modified one was SRB87, in which the author implemented overlapping PCR technique to develop a mutant strain devoid of stem-loop terminator structure responsible for leucine-dependent transcriptional regulation. These strains were converted into surfactin producer by the introduction of *sfp* gene of *B. subtilis* ATCC 21332 in the *amyE* locus by homologous recombination using pBG129 plasmid [3]. The strains were specified as *B. subtilis* BBG251 and BBG252, respectively. All the strains used in this paper are listed in Supporting information, Table S1.

2.1.3 Culture conditions and analysis

Each mutant obtained was cultivated in high-throughput system of fermentation Biolector® (Mp2-labs GmbH; Baesweiler; Germany) available on REALCAT platform of the University Lille 1. Cultures were carried out in 48 wells flower plate designed for the Biolector® (containing pH, dissolved oxygen and biomass probes). Cultures were performed at 37°C, under shaking (1100 rpm). Landy medium with 3.6 g/L (NH₄)₂SO₄ and TSS medium supplemented by 16 amino acids were used in these experiments [24]. These media were supplemented with tryptophan at 16 mg/L and buffered at pH 7.0 with MOPS 0.1 M. Samples were withdrawn after 6 h of culture. Biomass measurement for Biolector® calibration and surfactin production were carried out as described previously (see section 2.1.1). Experiments were carried out in triplicate and two biological replicates were performed. Results are expressed as mean value with standard deviation and statistical analysis of these data is made using one-way

ANOVA using a Student-Newman-Keuls post hoc test (p -value = 0.05). Different letters are used to indicate significant differences between the groups.

2.2 A modeling language for reaction networks

We first recall the modeling language for reaction networks with partial kinetic information from [20], and illustrate it at the regulation network of the promoter $P_{ilv-Leu}$, which is central for leucine production.

2.2.1 Graphical syntax

Most aspects of modeling language for reaction networks from [20] are quite standard in systems biology. The exceptions are the treatment of partially known kinetic functions and also for the treatment of inflow and outflow from the context.

A reaction network with partial kinetic knowledge consists of a finite set of species, a subset of inflow species, a subset of outflow species, and a finite set of reactions. A reaction is a multiset of pairs between species and roles. In the present paper, we fix the following set of roles: reactant, product, activator, accelerator, or inhibitor. A species is called a modifier of a reaction if its role is equal to activator, accelerator, or inhibitor. Modifiers are neither consumed nor produced, but the presence of a modifier affects the rate of the reaction. Note that the same species can be used several times in the same reaction and with different roles. The partial knowledge on the kinetics of a reaction is given by the roles that are assigned to its species. The intuition is as follows. The rate of a reaction is increased by reactants, activators, and accelerators, and it is decreased by inhibitors. A reaction is enabled only if all its reactants and activators are present, i.e. have nonzero concentrations, while some or the accelerators may be absent. We next explain the steady state semantics of a reaction network with partial kinetic information informally, before giving a formal definition in section 2.2.3.

As usual in deterministic semantics, we consider chemical solutions as functions from species to nonnegative real numbers including zero. For most kinds of species, this number represents the concentration of the species, but in other cases it may also denote the activity of a promoter or the activity of a promoter binding site. Reactions can be applied to chemical solutions. In this case, its reactants are consumed and its products are produced, respectively as many times as they occur in the reaction. A reaction can only be applied if all its activators are present (have a nonzero concentration), while its accelerators may be absent. The higher the concentration of an accelerator or activators, the higher will be the rate of the reaction. A reaction is applicable even if some inhibitors are present, but the higher the concentration the slower will be the rate of the reaction.

Reaction networks are usually situated in a context, which may be adjacent reaction networks, or the “chem-

ical medium” from which some species may inflow and to which some species may outflow during wet lab experimentation. The exchange with the context is modeled by the set of inflow and outflow species of the network: inflow species may inflow from the context, while outflow species may outflow into the context. The precise rate laws of inflows and outflows are unknown, except that we assume that the outflow rate must increase with the concentration of the outflow species.

Reaction networks in our modeling language are represented as graphs similar to Petri nets (the concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm). These graphs contain two kinds of nodes, round nodes for representing its species and boxed nodes for representing its reactions. More precisely, any species S is represented by a round node \circ and any reaction with name r by a boxed node \square . Solid edges either link a reactant to its reaction $\circ \rightarrow \square$, or a reaction to one of its products $\square \rightarrow \circ$. There are three kind of dashed edges, which start at the three kinds of modifiers. An accelerator edge links an accelerator to a reaction $\circ \circ \square$, an activator edge links an activator to a reaction $\circ \bullet \square$, and an inhibitor edge links an inhibitor to a reaction $\circ \dashv \square$. Whenever a species plays the same role several times in the same reaction, the multiplicity will be annotated to the corresponding edge. But this will not be the case in any of our examples.

An input edge $\circ \leftarrow$ points from the context to an inflow species S , while an outflow edge $\circ \rightarrow$ points from an outflow species S to the context. For convenience, we have introduced a last kind of edges $\square \rightarrow \circ$ as a shortcut for a product that is degraded by a hidden reaction, i.e. as a shortcut for: $\square \rightarrow \circ \rightarrow \square$.

An example network is given in Fig. 2. There we use species nodes with three different colors, which indicate their biological roles. The dynamics of the network is not concerned by the colors though. In yellow, we draw metabolites such as Leu and in blue proteins such as $CodY$. There is a third color for “artificial species” that serve for modeling regulation, such as for instance the promoter of the *ilv-leu* operon $P_{ilv-Leu}$. The real number that a chemical solution will assign to this artificial species will correspond to the activity of promoter $P_{ilv-Leu}$ in the solution (and not to its concentration).

In our graphs, we will annotate in orange, reactions that are potential candidates for knockouts. Dark orange indicates candidates that have been selected by our knockout prediction, while light orange indicates candidates that have not. We note again that node colors are irrelevant for the dynamics of the network.

2.2.2 Modeling the Regulation of $P_{ilv-Leu}$ Promoter

The *ilv-leu* operon contains the main genes involved in the production of branched chain amino acids and specially the leucine as shown in simplified manner in Fig. 2.

The *ilv-leu* operon is subject to multiple forms of regulation at its promoter $P_{ilv-Leu}$, as studied in a series of articles. Here we recall the model from [20] which considers the inhibition and activation events of the promoter, without explaining in detail how they are raised by binding to different promoter sites. The same abstraction was adopted earlier in one of the models for leucine production from the SubtiWiki [15], but these models lack a formal semantics as required for reasoning approaches.

We use the species Leu modeling the concentration of the metabolite leucine that we want to overproduce and let outflow to the context. We also introduce the species $P_{ilv-Leu}$ modeling the activity of the *ilv-leu* promoter. Reaction r_9 states that the expression of $P_{ilv-Leu}$ will finally lead to the expression of Leu in any case. This is an over-simplification adopted by this model, whose purpose is precisely to ignore the complex metabolic network behind the leucine production and its regulation. However, the promoter $P_{ilv-Leu}$ is also down-regulated by Leu in terms of ribosome-mediated attenuation mechanism T-box [24, 31], which is captured by that Leu is an inhibitor of r_2 , which activates $P_{ilv-Leu}$.

The regulation of $P_{ilv-Leu}$ is modeled by reaction r_2 that activates it. The degree of activation can also be degraded by the implicit reaction $r_{2'}$. Following [22, 24, 32], there are two proteins that down-regulate the activity of $P_{ilv-Leu}$ by binding to different sites upstream the promoter and inhibiting the RNA polymerase [33]. These are CodY (transcriptional pleiotropic regulator, which regulates the transcription of early-stationary-phase genes [34]) and TnrA (nitrogen pleiotropic transcriptional regulator), that we model by the actors BS_{CodY} and BS_{TnrA} respectively. The values of the corresponding variables $z_{BS_{CodY}}$ and $z_{BS_{TnrA}}$ represent the activity degrees of these binding sites. It can be increased by applying reactions r_1 and r_7 respectively, and decreased by applying the implicit degradation reactions $r_{1'}$ and $r_{7'}$. The degradation reactions r_{51} and r_{52} are necessary to balance the constant input fluxes of CodY or TnrA in steady states.

There is a further protein CcpA (Carbon catabolite control protein A), which is expressed by reaction r_{14} and can be degraded by the implicit reaction $r_{14'}$. Protein CcpA can bind to a site BS_{CcpA} of $P_{ilv-Leu}$ by reaction r_9 . This binding site overlaps with that of CodY [21]. Therefore, both proteins cannot bind simultaneously to the promoter, which we model by letting the reaction r_1 that binds CodY be subject to inhibition by CcpA. In addition, the binding of CcpA to the promoter $P_{ilv-Leu}$ accelerates the promoter [21, 22, 32]. Therefore, BS_{CcpA} is an accelerator of reaction r_2 . The catabolite acceleration by BS_{CcpA} can be inhibited when BS_{TnrA} is active, since then the active binding sites BS_{TnrA} and BS_{CcpA} can bind to each other while forming a DNA loop, which inhibits the acceleration. Therefore, BS_{TnrA} inhibits reaction r_9 .

The reactions of the network that can be disabled by gene knockouts are drawn in orange color. These are all

reactions modeling gene expressions, that are r_8 and r_{14} , and the reactions modeling promoter binding of the inhibitors, that is r_1 and r_7 .

2.2.3 Steady-state equations

We are interested in the deterministic steady state semantics of a reaction network, but projected to variables for exchange fluxes. Let \mathbb{R}_+ be the set of nonnegative real number including zero. For networks with n inflow species and m outflow species, the deterministic semantics of the network will be an exchange relation in $(\mathbb{R}_+)^{n+m}$. In our example, we have $n = 2$ and $m = 1$. However, it is never possible to determine the steady state semantics completely, given that we have only partial kinetic knowledge. Furthermore, note that we don't know about any initial concentrations, so it may be possible that different steady states are reached with different initial concentrations, even if we had complete kinetic knowledge otherwise.

In addition, we are usually given a set of k reactions that are knockout candidates, but do not know which of them will be knocked out. In our example, we had $k = 4$ knockout candidates. In this case, the steady state semantics will be enriched with unknowns for knockout candidates, and the projection will be to variables for both knockout candidates and exchange fluxes. We thus obtain a richer knockout-exchange relation in $(\mathbb{R}_+)^{k+n+m}$.

Given a species s of the network, we will introduce a variable z_s for its concentration in a steady state. For any inflow species we will introduce an influx variable x_s for the rate of the inflow of S , and for any outflow species we will introduce an outflux variable y_s for the rate of the outflow of S . Furthermore, we will use variables v_{r_i} for the speed of reaction r_i in a steady state, and 0/1 variables o_{r_i} ; stating whether reaction r_i is off or on respectively.

One possible kinetic functions $kin_{S,Act,Acc,I}$ for a reaction, with a set of substrates S , activators Act , accelerators Acc and inhibitors I is equal to:

$$kin_{S,Act,Acc,I}(z_s)_{s \in S, Act, Acc, I} = \prod_{s \in S} z_s \prod_{a \in Act} z_a \frac{(1 + \sum_{a \in Acc} z_a)}{1 + \sum_{i \in I} z_i}$$

The function $kin_{S,Act,Acc,I}$ satisfies the basic requirements, which beside its well definedness for all possible concentrations are that:

- it is monotonic in the concentrations of all substrates, activators, and accelerators, and
- it is anti-monotonic in the concentrations of all inhibitors, and
- it is 0 if the concentration of one of the activators or substrates becomes 0.

However, all the constants in the definition of $kin_{S,Act,Acc,I}$ are equal to 1. This is clearly randomly chosen, so there is no reason to believe that $kin_{S,Act,Acc,I}$ will be to the true

kinetic function of any chemical reaction in our example or elsewhere. In order to face this situation, we will assume that the kinetic functions of reactions with substrates S , activators Act , accelerators Acc , and inhibitors I are similar to $kin_{S,Act,Acc,I}$. The precise definition of similarity in Definition 1 will be such that the above properties are satisfied.

So let $kin^{(i)}$ be a variable for the kinetic function of reaction r_i . Any steady state of the $P_{Iiv-Leu}$ regulation network must then satisfy the equations below. The first equation states that the speed v_{r_i} of reaction r_i is equal to 0 if the reaction is knocked out, so that $o_{r_i} = 0$, and equal to $kin^{(1)}(i : z_{CcpA}, act : z_{CodY})$ otherwise. In addition, the model imposes the constraint that unknown function $kin^{(1)}$ is similar to the known function $kin_{\emptyset,(CodY),\emptyset,(CcpA)}$. The equations for the speeds of all other reactions are similar. Note that the roles of arguments are always indicated by attributes (i , inhibitor in I ; act , activator in Act ; acc , accelerator in Acc ; s : substrate in S).

Steady state equations

Reactions speeds

$$\begin{aligned} v_{r_1} &= o_{r_1} kin^{(1)}(i : z_{CcpA}, act : z_{CodY}) \\ v_{r_2} &= kin^{(1)}(s : z_{BS_{CodY}}) \\ v_{r_3} &= kin^{(2)}(i : z_{BS_{CodY}}, acc : z_{BS_{CcpA}}, i : z_{Leu}, i : z_{BS_{TtrA}}) \\ v_{r_4} &= kin^{(2)}(s : z_{P_{Iiv-Leu}}) \\ v_{r_5} &= o_{r_5} kin^{(7)}(act : z_{TtrA}) \\ v_{r_6} &= kin^{(7)}(s : z_{BS_{TtrA}}) \\ v_{r_7} &= kin^{(9)}(act : z_{CcpA}, i : z_{BS_{TtrA}}) \\ v_{r_8} &= kin^{(9)}(s : z_{BS_{CcpA}}) \\ v_{r_9} &= kin^{(5)}(s : z_{CodY}) \\ v_{r_{10}} &= kin^{(52)}(s : z_{TtrA}) \\ v_{r_{11}} &= o_{r_{11}} kin^{(14)}() \\ v_{r_{12}} &= kin^{(14)}(s : z_{CcpA}) \\ v_{r_{13}} &= o_{r_{13}} kin^{(8)}(s : z_{P_{Iiv-Leu}}) \\ v_{r_{14}} &= kin^{(0)}(s : z_{Leu}) \end{aligned}$$

Flux balance equations

$$\begin{aligned} y_{Leu} &= v_{r_8} \\ v_{r_{12}} &= v_{r_{11}} \\ v_{r_{10}} &= x_{CodY} \\ v_{r_{13}} &= x_{TtrA} \\ v_{r_1} &= v_{r_1} \\ v_{r_2} + v_{r_3} &= v_{r_2} \\ v_{r_6} &= v_{r_6} \\ v_{r_7} &= v_{r_7} \end{aligned}$$

Outflow model

$$y_{Leu} = kin^{(0)}(z_{Leu})$$

In a (limit) steady state, we need that all species are consumed and produced at equilibrium. For $CcpA$ for instance, this requires that $v_{r_{12}} = v_{r_{11}}$ and for $CodY$ that $v_{r_{10}} = x_{CodY}$. Finally, we have to impose that an outflux increases monotonically with the concentrations of the outflow species. For Leu this means that $y_{Leu} = kin^{(0)}(s : z_{Leu})$ for some kinetic function $kin^{(0)}$ that is similar to $kin_{(Leu),\emptyset,\emptyset,\emptyset}$. Without such an outflow model, we would lack any relation between the concentration of the species z_{Leu} and the speed of its outflux y_{Leu} . This would spoil our knockout predictions for leucine overproduction in particular.

The flux balance equations can be used to simplify the remaining equations by replacing equal by equal, while removing some local variables (those for reaction speeds v_{r_i} and species concentrations z_j). The simplified equations are given below. It should be noticed that such simplifications before abstract interpretation improve the precision of this analysis.

Simplified steady state equations

$$\begin{aligned} v_{r_1} &= o_{r_1} kin^{(1)}(i : z_{CcpA}, act : z_{CodY}) \\ v_{r_1} &= kin^{(1)}(s : z_{BS_{CodY}}) \\ v_{r_3} &= kin^{(2)}(i : z_{BS_{CodY}}, acc : z_{BS_{CcpA}}, i : z_{Leu}, i : z_{BS_{TtrA}}) \\ v_{r_4} &= kin^{(2)}(s : z_{P_{Iiv-Leu}}) + v_{r_8} \\ v_{r_5} &= o_{r_5} kin^{(7)}(act : z_{TtrA}) \\ v_{r_6} &= kin^{(7)}(s : z_{BS_{TtrA}}) \\ v_{r_7} &= kin^{(9)}(act : z_{CcpA}, i : z_{BS_{TtrA}}) \\ v_{r_8} &= kin^{(9)}(s : z_{BS_{CcpA}}) \\ x_{CodY} &= kin^{(5)}(s : z_{CodY}) \\ x_{TtrA} &= kin^{(52)}(s : z_{TtrA}) \\ v_{r_{10}} &= o_{r_{10}} kin^{(14)}() \\ v_{r_{11}} &= kin^{(14)}(s : z_{CcpA}) \\ y_{Leu} &= o_{r_{13}} kin^{(8)}(s : z_{P_{Iiv-Leu}}) \\ y_{Leu} &= kin^{(0)}(s : z_{Leu}) \end{aligned}$$

2.2.4 Similarity

When doing flux analysis or knockout prediction, we are interested in what may happen with the steady states, when the model is changed, either by changing some influxes or outfluxes, or by knocking out some of the reactions, or both.

The state space for the change of concentrations or of a reaction speed $\mathbb{R}_+ \times \mathbb{R}_+$. We partition this space into the six difference relations in $\Delta = \{\uparrow, \downarrow, \uparrow\downarrow, \downarrow\uparrow, \sim, \approx\}$. The increase relation $\uparrow = \{(x, y) \mid 0 < y\}$ contains all increases but not from zero, while the increase-from-zero relation $\uparrow\downarrow = \{(0, y) \mid 0 < y\}$ contains all other increases. The relation \downarrow captures all decreases but not to zero, while $\downarrow\uparrow$ covers all decreases to zero. There are also two ways of no-change, those at a value different to zero $\sim = \{(x, x) \mid x \neq 0\}$ and the

no-change at zero $\approx = \{(0,0)\}$. Let R be a relation in $(\mathbb{R}_+)^p$ for some $p \geq 0$. We define the abstract interpretation of R over Δ as follows:

$$R^\Delta = \left\{ \left(\delta_1, \dots, \delta_p \right) \in \Delta^p \mid \exists (x_1, \dots, x_p) \in R, (y_1, \dots, y_p) \in \delta_p \right\}$$

Since kinetic functions $\kappa: (\mathbb{R}_+)^{p-1} \rightarrow \mathbb{R}_+$ can be seen as p -ary relations, their abstract interpretations κ^Δ are well defined.

Definition 1. We call two kinetic functions $\kappa_1, \kappa_2: (\mathbb{R}_+)^{p-1} \rightarrow \mathbb{R}_+$ similar if and only if $\kappa_1^\Delta = \kappa_2^\Delta$.

It can now be verified that all kinetic functions similar to $kin_{S,Act,Acc,I}$ satisfy the three properties required above. Furthermore, similarity is preserved when changing the parameter of mass action kinetics. It can even be seen, that the binary mass action kinetics is similar to the Michaelis–Menten kinetics.

3 Results

3.1 Knockout prediction algorithm

We now generalize the knockout prediction algorithm from [19] so that it applies to all reaction networks modeled in the language recalled in section 2.2. The difficulty compared to there is to overcome the limitation to reactions with mass action kinetics. This will be done by extending the reasoning methods from [20] for our modeling language so that we can predict reaction knockouts, instead of changes of influxes or outfluxes only. We have already prepared this extension by adding the following two aspects from [19] to our presentation of the modeling language from [20]. Firstly, we insert knockout variables o_i into steady state equations in section 2.2.3. And secondly, we have chosen the domain with the six elements $\Delta = \{\uparrow, \downarrow, \uparrow, \downarrow, \approx, \approx\}$ for abstract interpretation in section 2.2.4, rather than the three element domain $\Delta' = \{<, =, >\}$. The intuition is that the change performed by the knockout of a reaction r_i can be reflected by a difference constraint $o_{r_i} = \downarrow$, meaning that r_i was on before and off after the change.

3.1.1 Abstract interpretation

Abstract interpretation can be lifted to the steady state equations as obtained from a reaction network. For this, all variables get interpreted over Δ (instead of \mathbb{R}_+), and all arithmetic functions are interpreted abstractly over Δ (instead of \mathbb{R}_+) as defined in section 2.2.4: Addition is interpreted as $+$, multiplication as \cdot , and unknown kinetic functions $kin^{(i)}$ as the fully known function $kin^{(i)} = (kin_{S,Act,Acc,I})^\Delta$. It should be noticed that functions can become set valued by abstract interpretation. For

instance, $+^\Delta(\uparrow, \downarrow) = \{\uparrow, \downarrow, \approx\}$, so that abstract interpretation will produce difference constraints of the form $z_1 \in z_2 + z_3$.

The abstract interpretations of the simplified steady state equations, for instance, are given in below.

Difference constraints obtained by abstract interpretation, where:

$$inh(\mathbf{x}) = \frac{1}{acc(\mathbf{x})} \text{ and } acc(\mathbf{x}) = 1 + \mathbf{x}$$

$$\begin{aligned} V_{r_1} &\in O_{r_1} \cdot inh(z_{CcpA}) \cdot z_{CodY} \\ V_{r_1} &= z_{BS_{CodY}} \\ V_{r_2} &\in inh(z_{BS_{CodY}} + z_{Leu} + z_{BS_{TnrA}}) \cdot acc(z_{BS_{CcpA}}) \\ V_{r_2} &\in z_{P_{Tr-Leu}} + V_{r_2} \\ V_{r_7} &\in O_{r_7} \cdot z_{TnrA} \\ V_{r_7} &= z_{BS_{TnrA}} \\ V_{r_9} &\in z_{CcpA} \cdot inh(z_{BS_{TnrA}}) \\ V_{r_9} &= z_{BS_{CcpA}} \\ X_{CodY} &= z_{CodY} \\ X_{TnrA} &= z_{TnrA} \\ V_{r_{14}} &= O_{r_{14}} \\ V_{r_{14}} &= z_{CcpA} \\ Y_{Leu} &\in O_{r_9} \cdot z_{P_{Trv-Leu}} \\ Y_{Leu} &= z_{Leu} \end{aligned}$$

Most importantly, we could replace the variables $kin^{(i)}$ by the definition of $kin_{S,Act,Acc,I}$ since $(kin^{(i)})^\Delta = (kin_{S,Act,Acc,I})^\Delta$. In this way, we successfully abstracted away the lacking kinetic knowledge.

We next recall the main insight from John et al. [19], which states that the difference assignment between two steady states of a given the reaction network must satisfy the abstract interpretation of its steady state equations. Indeed, the proposition is slightly more general, so that it remains independent of the choice of the modeling language. For any two variable assignments v, v' into \mathbb{R}_+ and any variable x they consider the unique difference relation $\delta(x) \in \Delta$ such that $(v(x), v'(x)) \in \delta(x)$.

Proposition 1 (John et al. [19]). *If v and v' are two solutions of the arithmetic equations \mathcal{O} over \mathbb{R}_+ , then their difference assignment δ satisfies the abstract interpretation of \mathcal{O} over Δ .*

The solutions v and v' of interest are the steady states of a reaction network with simplified steady state equations \mathcal{O} . More precisely, v should be a solution of \mathcal{O} without any knockout and v' another solution of \mathcal{O} with some knockout. The abstract interpretation of \mathcal{O} must then be satisfied by all possible changes of δ from v to v' .

3.1.2 Knockout prediction by abstract interpretation

A knockout prediction problem receives as inputs a reaction network, a subset of reactions that are candidates for knockouts, and an overproduction target. In the case of

Table 1.

A: The top-4 solutions of the difference constraints and the overproduction target $y_{\text{Leu}} = \uparrow$

solution	o_{r_1}	o_{r_7}	o_{r_8}	$o_{r_{14}}$	x_{CodY}	x_{TnrA}	y_{Leu}	penalty
1.	~	~	~	~	↓	~	↑	2
2.	~	~	~	~	~	↓	↑	2
3.	↓	~	~	~	~	~	↑	2
4.	~	↓	~	~	~	~	↑	2

B: Top-2 solutions: no knockouts and 4 context changes.

solution	x_{Cltu}	x_{Pyr}	$y_{\text{Acyl-CoA}}$	y_{Leu}	y_{OxoGlu}	penalty
1.	↑	↑	~	↑	↑	4
2.	↑	~	↓	↑	↑	4

For any i and S , $o_i \downarrow$ means a knockout of reaction r_i , $x_S \downarrow$ a decrease of the influx of species S (but not to zero), and $y_S \uparrow$ an increase of the outflux of S (but not from zero).

the $P_{\text{Ilv-Leu}}$ network, the knockout candidates are r_1 , r_7 , r_8 , and r_{14} , and the overproduction target is $y_{\text{Leu}} = \uparrow$. The question then is which of the knockout candidates can be knocked out in the network, while satisfying the overproduction target.

Let \emptyset be the simplified steady state equations of the network. The question then is whether there exists two variable assignments v and v' solving \emptyset such that $v(o_{r_i}) = 1$ for all knockout variables in \emptyset , and such that the difference assignment of v and v' satisfies the overproduction target. In this case, a solution for the knockout problem is the subset of all reactions r_i such that $v(o_{r_i}) = 0$.

In order to search for two variable assignments satisfying \emptyset and the overproduction target, we can apply Proposition 1. It shows that it is sufficient to search for a variable assignment satisfying the abstract interpretation of \emptyset over Δ . Furthermore, we require to be satisfied the overproduction target, the constraints $o_{r_i} = \{\downarrow, \sim\}$ for all reactions r_i that can be knocked out, and $o_{r_i} = \{\sim\}$ for all others, and the flux restrictions $x_S y_S \in \{\uparrow, \downarrow, \sim\}$ for all species S .

The number of solutions is often huge. Therefore, we are interested only in those solutions where the overall number of knockouts, influx changes, and outflux changes is as small as possible. The hope is that the solutions with the minimal number of changes contain exactly those changes that are really needed to satisfy the overproduction target. In order to restrict the search space further, we will rule out all solutions in $o_{r_i} = \downarrow$ for more than one i .

3.2 Application to $P_{\text{Ilv-Leu}}$ example network

We define the penalty of a solution as the number of global variables mapped to \uparrow, \downarrow , or \sim . As argued above, the lower is the penalty; the better is the solution, since no irrelevant changes were added besides those that are really needed.

The top- n solutions of a difference constraint can be computed by a finite domain constraint solver with branch and bound optimization. The top-4 solutions of the difference constraints for the $P_{\text{Ilv-Leu}}$ network are given in Table 1A; these are exactly all solutions with penalty 2. Of course, there are many further solutions with higher penalties, but the top-4 solutions indeed correspond to the most interesting predictions in this case:

1. Decrease influx of CodY, or
2. Decrease influx of TnrA, or
3. Knockout reaction r_1 , or
4. Knockout reaction r_7 .


What are predicted for this small network are two flux changes and two single knockouts. However, a decrease of the fluxes x_{CodY} and x_{TnrA} can be obtained by knocking out the respective genes in the context, i.e. reaction r_{15} and r_{16} in the larger leucine network in Fig. 3. Therefore, the four predictions for the small $P_{\text{Ilv-Leu}}$ network correspond to four single knockout predictions for the larger leucine model. This illustrates the modularity of our approach. It is obtained by admitting inflow and outflow species for the interaction of a module of a network with its context, without need to model the whole network.




3.3 Modeling of leucine metabolic pathway

A larger and more realistic model of leucine production is presented in Fig. 3. It describes the metabolic transformations of Thr, Pyr and Akb into Ile, Val, and Leu, concomitantly with its regulation. Our model mainly follows the abstraction level chosen by the previous work in [15, 26], but the regulation is expressed more cleanly within our formal modeling language. In the Supporting information, we list the metabolites, proteins and actors of this model in Table S3, and all its reactions in Table S4.

Generally, we will assume that the precise kinetics are unknown, but hope that we still have good knowledge on inhibitors and activators of the reactions. We do not have to know all inhibitors and activators, but the more we

know the better predictions we can expect. This is an important departure from previous methods purely based on flux balance equations.

In order to group several proteins that all participate in the same reaction, we introduce artificial species that we call protein clusters, and draw them as for instance . This protein cluster presents all the proteins generated by the *bkL* operon.

We will slightly enrich our graphical syntax with some syntactic sugar, in order to keep bigger models readable. This sugar can be compiled away in a preprocessing step. First, we use split-points  that link different copies of the same species. For instance, we have one copy of Leu, which are drawn as  and linked together with the original  over a split point. Logically, there is no difference between a species and its copies; it just allows nicer graphs with fewer intersections of edges. Second, we use edge-clusters that group several with the same sources or targets (for instance the activator edges from BkL and YwaA to r_{35} , r_{32} , and r_{29} are clustered).

3.3.1 Production of Isoleucine, Valine, and Leucine

For clarity, we here only consider those intermediates that produce the outflows of the pathway, i.e. Ket_a for Ile in r_{26} , Ket_b for Val in r_{36} , and Ket_c for Leu in r_{45} . All three reactions must be activated by YwaA+YbgE, that is either by protein YwaA (r_6) or protein YbgE (r_7) which are exchangeable catalysts following [35]. There are also inverse reactions r_{29} , r_{32} , r_{35} , but with different activators, the proteins Bcd and YwaA. The intermediates Ket_a , Ket_b , and Ket_c can be transformed alternatively into Acyl-CoA by reactions r_{28} , r_{31} , and r_{34} , which can then flow out into the context for fatty acid biosynthesis. This is regulated by protein BkL.

The creation of Ile, Val, and Leu from Ket_a , Ket_b , Ket_c is always coupled with a parallel transformation of Glu into OxoGlu, and the inverse decompositions are coupled with the inverse transformation of OxoGlu into Glu. The required Glu for the production of Ile, Val, and Leu, can be input from the glutamate biosynthesis cycle of the context (derived from OxoGlu) produced OxoGlu can be flow out from this pathway to the context (as for example used in the TCA cycle). The intermediates are produced as follows: Ket_a is produced from Akb by r_{27} , Ket_b from Pyr by r_{30} , and Ket_c from Ket_b by r_{33} . Akb can be produced from Thr by reaction r_{41} , which in turn can also be input from the threonine biosynthesis pathway of the context (which is linked to the glycolysis pathway via the aspartate biosynthesis). Pyr can be input from the glycolysis pathway of the context. IlvA activates the conversion of Thr into Akb. IlvBH (a cluster of IlvB and IlvH), IlvC and IlvD activate the transformation of Akb to Ket_a . The proteins LeuA and LeuBCD (a cluster of LeuA, LeuB, and LeuC) perform a sequence of reactions justifying the transformation of Ket_b to Ket_c [26]. The genes of the proteins IlvBH, IlvC, LeuA, and LeuBCD are co-located in the same

operon (*ilv-leu*), so they are under the dependence of one and the same promoter $P_{\text{Ilv-Leu}}$ [22]. This specific regulation is detailed below.

3.3.2 Regulation

The expression of the different enzymes involved in the production of branched chain amino acids (explained above) are denoted by the following reactions: r_{22} for IlvA; r_{25} for IlvD; r_{23} for IlvBH; r_{24} for IlvC; r_{38} for LeuBCD; r_{37} for LeuA; r_{11} for BkL; r_{10} for Bcd; r_{47} for YwaA and r_{46} for YbgE. All these expression reactions are controlled by a complex regulation [24, 36, 37]. Indeed, we consider proteins CcpA, CodY, TnrA, and BkdR that have regulatory functions and are expressed respectively by the reactions r_{14} , r_{15} , r_{16} and r_{12} , r_{13} . The expression reactions of these regulators can be modulated by different factors from outside the context, which are not modeled here. Gtp comes from the context and can be degraded through the reaction r_{17} . This metabolite increases the affinity of CodY for its binding sites, as well as Ile and Val [21–23, 36, 37].

Expression of IlvA, IlvD, and YbgE is dependent on their own promoters and is down-regulated by CodY at the transcriptional level [36, 37], which is represented by reactions r_{22} , r_{25} , and r_{46} , respectively. In [38] it was further shown that IlvA is deactivated by both Ile and Val (reactions r_{20} , r_{21}). The transcription of all proteins IlvC, IlvBH, LeuA, and LeuBCD starts through the activation of their promoter $P_{\text{Ilv-Leu}}$, which controls the reactions r_{23} , r_{24} , r_{37} and r_{38} , respectively [24, 37]. Its regulation is captured by the reaction r_2 and is dependent on several mechanisms, which are described before in section 2.2.2. In addition to what has been described above we introduced here the down-regulation of the promoter $P_{\text{Ilv-Leu}}$ by Leu in terms of ribosome-mediated attenuation mechanism Tbox (r_{40}) [24, 31] and the deactivation of LeuA by Leu which is represented by the reaction r_{39} . Proteins BkL and Bcd lie under the action of two common but independent promoters (Prm_b and Prm_c), whose activity level is represented by actor $\text{OP}_{\text{BkL-Bcd}}$. Among them one is constitutive (no regulation, represented by the reaction r_4), and the other is positively impacted by BkdR and down-regulated by both TnrA and CodY, as reflected by reaction r_3 [39]. Although not regulated on the transcriptional level, protein BkdR is activated by Ile and Val, which is modeled by introducing the two BkdR producing reactions r_{12} and r_{13} . Proteins YbgE (expressed by the reaction r_{46}) is down regulated by CodY. However, no regulation is known for protein YwaA, which is expressed by the reaction r_{47} .

3.4 Knockout prediction for leucine overproduction

As for the small model, our target for knockout prediction is leucine overproduction, i.e. $y_{\text{Leu}} = \uparrow$. But in the big model, there are also some unwanted possibilities to do so, which are to decrease the outfluxes of valine or isoleucine, since the remaining valine and isoleucine could then be

Table 2. Next best solutions: a single knockout and four context changes.

solution	knockouts							context changes					penalty
	o_{r_1}	$o_{r_{10}}$	$o_{r_{12}}$	$o_{r_{13}}$	o_{r_3}	o_{r_4}	o_{r_7}	x_{Glu}	x_{Pyr}	$y_{Acyl-CoA}$	y_{Leu}	y_{OxoGlu}	
3.	↓	~	~	~	~	~	~	↑	↑	~	↑	↑	5
4.	↓	~	~	~	~	~	~	↑	~	↓	↑	↑	5
5.	~	~	~	~	↓	~	~	↑	↑	~	↑	↑	5
6.	~	~	~	~	↓	~	~	↑	~	↓	↑	↑	5
7.	~	~	~	~	~	↓	~	↑	↑	~	↑	↑	5
8.	~	~	~	~	~	↓	~	↑	↑	↓	↑	↑	5
9.	~	~	~	~	~	~	↓	↑	↑	~	↑	↑	5
10.	~	~	~	~	~	~	↓	↑	~	↓	↑	↑	5
11.	~	↓	~	~	~	~	~	↑	↑	~	↑	↑	5
12.	~	↓	~	~	~	~	~	↑	~	↓	↑	↑	5
13.	~	~	↓	~	~	~	~	↑	↑	~	↑	↑	5
14.	~	~	↓	~	~	~	~	↑	~	↓	↑	↑	5
15.	~	~	~	↓	~	~	~	↑	↑	~	↑	↑	5
16.	~	~	~	↓	~	~	~	↑	~	↓	↑	↑	5

solution	knockouts						context changes					penalty
	$o_{r_{47}}$	$o_{r_{13}}$	$o_{r_{16}}$	$o_{r_{40}}$	$o_{r_{46}}$	$o_{r_{14}}$	x_{Glu}	x_{Pyr}	$y_{Acyl-CoA}$	y_{Leu}	y_{OxoGlu}	
17.	~	↓	~	~	~	~	↑	↑	~	↑	↑	5
18.	~	↓	~	~	~	~	↑	~	↓	↑	↑	5
19.	~	~	↓	~	~	~	↑	↑	~	↑	↑	5
20.	~	~	↓	~	~	~	↑	~	↓	↑	↑	5
21.	~	~	~	↓	~	~	↑	↑	~	↑	↑	5
22.	~	~	~	↓	~	~	↑	~	↓	↑	↑	5
23.	↓	~	~	~	~	~	↑	↑	~	↑	↑	5
24.	↓	~	~	~	~	~	↑	~	↓	↑	↑	5
25.	~	~	~	~	↓	~	↑	↑	~	↑	↑	5
26.	~	~	~	~	↓	~	↑	~	↓	↑	↑	5
27.	~	~	~	~	~	↓	↑	↑	~	↑	↑	5
28.	~	~	~	~	~	↓	↑	~	↓	↑	↑	5

solution	knockouts	context changes						penalty
	$o_{r_{11}}$	x_{Glu}	x_{Thr}	$y_{Acyl-CoA}$	y_{Ile}	y_{Leu}	y_{OxoGlu}	
142.	↓	↑	↑	↓	~	↑	↑	6
143.	↓	↑	~	↓	↑	↑	↑	6

For any i and S , $o_i \downarrow =$ means a knockout of reaction r_i , $x_S \downarrow =$ a decrease of the influx of species S (but not to zero), and $y_S \uparrow =$ an increase of the outflux of S (but not from zero).

used to produce more leucine. In order to rule out this, we add to our knockout prediction target that the outfluxes of Val and Ile cannot decrease, i.e. $y_{Val}, y_{Ile} \in \{~, \uparrow\}$.

We first present the solutions of the solver for difference constraints, and then show how they can be interpreted as knockout predictions. In all the solutions we observed it holds that $x_{Glu} = y_{OxoGlu} = y_{Leu} = \uparrow$, so this seems to be always the case. This shows that any increase of y_{Leu} requires increasing the influx of Glu and the outflux of OxoGlu. It implies for all solutions, that their minimal overall penalty is at least 3.

The top-2 solutions presented in Table 1B have penalty 4. These do not require any reaction knockout. Besides the three necessary context changes discussed above, we either have that $y_{Acyl-CoA} = \downarrow$, i.e. the outflux of Acyl-CoA is decreased, or else that $x_{Pyr} = \uparrow$, so that the influx of Pyr

is increased. Both possibilities are plausible, since for overproducing Leu, either the production of Leu by r_{45} must be increased, and this requires more Pyr, or else the outflux of Leu into the fatty acid pathway (via reactions r_{35} and r_{34}) must be decreased, and this imposes a higher outflux of Acyl-CoA.

All the single knockout solutions with penalty 5 are given in Table 2, with numbers 3–28. Indeed, it turns out that they all extend on either of the top-2 solutions. This is a rather particular and surprising situation. It means that none of these knockouts is strictly necessary to justify Leu overproduction, but that they are all compatible with it. In other words, all other candidates who were not selected are incompatible with a single knockout and 4 context changes. A further single knockout is found with penalty 6 in the solutions with number 142 and 143.

Table 3. Single knockout predictions for leucine overproduction.

knockout	description of reaction	knockout effect	mutant
r_1	CodY binding to BS_{CodY}	increases $P_{\text{Ilv-Leu}}$ activity	BBG251
r_3	binds BkdR to Bkl-Bcd promoter	decrease Bcd and thus r_{35}	
r_4	constitutive expression of Bkl-Bcd operon	decreases Bcd and thus r_{35}	
r_7	TnrA binding to BS_{TnrA}	increases $P_{\text{Ilv-Leu}}$ activity	
r_{10}	Bcd expression	deactivates reaction r_{35}	BBG253
r_{11}	Bkl expression	deactivates Ket_1 , Ket_2 , Ket_3 outflow via Acyl-CoA by r_{28} , r_{31} , r_{34}	BBG255
r_{12}	activates BkdR by Ile	decreases speed of r_3 and thus Bcd	BBG254
r_{13}	activates BkdR by Val	decreases speed of r_3 and thus Bcd	
r_{15}	CodY expression	increases $P_{\text{Ilv-Leu}}$ activity	
r_{16}	TnrA expression	increases $P_{\text{Ilv-Leu}}$ activity	
r_{40}	Tbox of Leu attenuation	increases $P_{\text{Ilv-Leu}}$ activity	BBG256
r_{47}	expression of YwaA	increases speed of r_{35}	BBG252
r_{14}	expression of CcpA	unclear	
r_{46}	expressions of YbgE	unclear	

There are 38 further solutions with a penalty of 5 which do not require any knockouts. Of these 12 are extensions of the top-2 solutions by further context changes, and the others remove assign one knockout variable to \approx , meaning that one reaction is removed from the model. There are many further solutions with penalty 6 ranging from 67 to 784 and many more solutions with penalty 7 ranging from 785 to 4249. But these do not add further single knockout predictions. The computation of all solutions up to penalty 7 requires about 6:21 minutes on a MacBook Pro laptop (2.4 GHz Intel core i7). For higher penalties, our algorithm is running out of memory.

All solutions corresponding to single knockouts are collected in Table 3. This can be done automatically by our tools. In addition, we have added an informal explanation of the positive effect of these single knockouts with respect to leucine overproduction. These effects can be seen in the model rather easily. It turns out, however that knockout of reactions r_{14} and r_{46} do not have clear positive effect. The only effect of a knockout of reaction r_{46} is a deactivation of r_{45} producing Leu, which is unwanted. The knockout of reaction r_{14} has a negative effect on the activity of $P_{\text{Ilv-Leu}}$. This is not very plausible for leucine overproduction: it could still be argued that a decrease of the activity of $P_{\text{Ilv-Leu}}$ will decrease the speed of r_{27} which in turn will increase the speed of r_{30} , but at the same time, it also has a negative effect on the activation of r_{30} , which counter balances this positive effect on the Leu production. Moreover, it has been demonstrated previously that a knockout of *ccpA* causes a decrease of *ilv-leu* expression [24].

3.5 Effect of leucine feeding in surfactin overproduction

In order to check the working hypothesis that an increase in the intracellular pool of leucine leads to an increase of

the surfactin production, a preliminary experiment was performed by adding leucine in the Landy medium (instead of $(\text{NH}_4)_2\text{SO}_4$). In the medium with only $(\text{NH}_4)_2\text{SO}_4$ as the nitrogen source the specific surfactin production reached 117.09 ± 19.62 mg/g of DW. In contrast, the presence of leucine in the medium led to three-fold increase of the specific surfactin production of BBG111 (328.41 ± 31.94 mg/g of DW) (data not shown). This experiment shows that an increase in the extracellular concentration of leucine certainly produces an intracellular increase, which was beneficial for the production of surfactin. This confirms that intracellular concentration of the precursors is one of the most limiting factors for surfactin overproduction. Therefore, the computational modeling of the metabolic pathway of leucine and the knockout prediction obtained appears to be consistent and can be checked by genetic engineering.

3.6 Genetic engineering of *B. subtilis* for the overproduction of surfactin and one of its precursor

In this part, we present the different phenotypic profiles of the mutant strains obtained through genetic engineering of *B. subtilis* in terms of growth and specific surfactin production. These results were obtained in different culture media when the cells were in the exponential growth phase (about 6 h of culture). Indeed, at this time the *srfA* operon's promoter reaches its maximum activity [3, 40], as well as the *ilvB* operon is highly expressed [26, 29].

3.6.1 Genetic engineering of *B. subtilis* and growth analysis

According to the prediction results based on the model, various genes were deleted. Firstly, to inhibit the CodY binding to its high affinity site on the $P_{\text{Ilv-Leu}}$ promoter (r_1) and secondly, to delete the *codY* expression (r_{15}), the

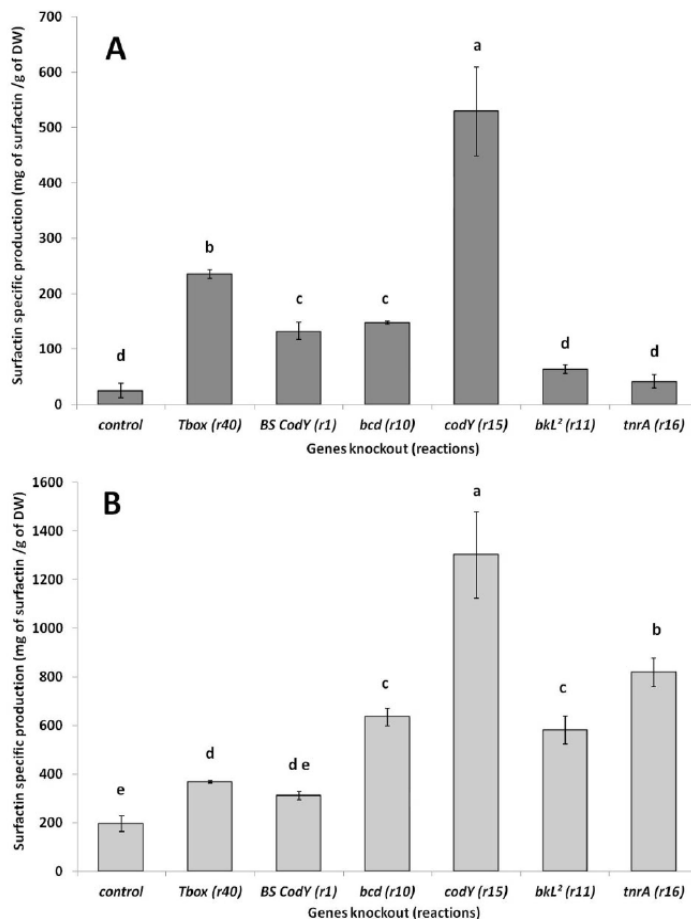


Figure 4. Specific surfactin production of the different knockout mutants strains after 6 h of culture in TSS medium containing a mix of 16 amino acids as nitrogen source (A) and in Landy medium containing $(\text{NH}_4)_2\text{SO}_4$ as nitrogen source (B). Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA with a Student-Newman-Keuls post hoc test (p -value = 0.05). Different letters were used to indicate significant differences between groups. ² Whenever we talk about a knockout for the *bkl* operon, we mean more precisely the *lpdV* gene of this operon.

strains BBG251 and BBG254 were constructed. TnrA also negatively regulates the *ilv-leu* expression under nitrogen-limited condition, the knockout predictions giving as the result the deletion of the TnrA binding on the $P_{\text{ilv-leu}}$ promoter (r_7) and the expression of *tnrA* (r_{16}). To achieve the same effect, we have chosen to completely suppress the expression of *tnrA* (r_{16}) in the mutant strain BBG256.

Leucine intracellular concentration can inhibit the expression of *ilv-leu* operon via Tbox attenuator. In the mutant strain BBG252, Tbox was suppressed (r_{40}), as recommended by the predictions. Lastly, predictions have suggested to avoid leucine degradation by the knockout of the reactions r_3 , r_4 , r_{12} , r_{13} and r_{47} . To get this, the expression of *bcd* (r_{10}) and *bkl* (r_{11}) were suppressed, giving the strains BBG253 and BBG255, respectively. Growth kinetics study was carried out during the exponential growth

phase in Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ and in TSS medium with 16 amino acids (see Supporting information, Table S5). The specific growth rate of the control strain BBG111 was found to be $0.47 \pm 0.06 \text{ h}^{-1}$ and $0.55 \pm 0.05 \text{ h}^{-1}$ in the former and later media respectively. The specific growth rates of various mutants were not significantly different from the control strain BBG111 in the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$. However in TSS medium with 16 amino acids, only the specific growth rates of BBG253 and BBG254 were lower and significantly different from that of the control strain.

3.6.2 Surfactin yield as an indicator for increase in intracellular leucine

Specific surfactin production for BBG111 along with the mutant strains is shown in Fig. 4A for TSS medium with

16 amino acids and in Fig. 4B for Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ and summarized in Supporting information, Table S5. In TSS medium with 16 amino acids, the specific surfactin production of the different mutant strains is better than the control strain. Statistical analysis reveals that the results obtained for BBG256 (*tnrA* deletion) and BBG255 (*bkL* deletion) mutants were not significantly different from the control strain. Moreover, the specific surfactin production of BBG251 (*BS_{CodY}* deletion) and BBG253 (*bcd* deletion) can be considered in the same statistical group. In Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, the specific surfactin production of the different mutant strains is also higher than the control strain. Statistical analysis reveals that these results were significantly different from those obtained with the control strain, except for the strain BBG251, in which the CodY high affinity binding site was deleted (*BS_{CodY}*). The specific surfactin production of this strain can be also considered as close to those obtained with the strain BBG252 (*Tbox* deletion). Moreover, the specific surfactin production of BBG253 (*bcd* deletion) and BBG255 (*bkL* deletion), which are the two strains modified in the branched chain amino acids degradation pathway can be considered in the same statistical group.

4 Discussion

The metabolic flux analysis is a central method of cell engineering to increase the productivity of host microorganisms such as bacteria, yeast and fungi. In this area, many fundamental studies have been conducted on the pathways regulations and metabolic flux analysis [41]. The use of bioinformatics in recent years helped to develop tools for reconstruction of metabolic pathways and genome scale metabolic modeling in order to overexpress metabolites, but most of them are only applied to primary metabolites and not on secondary metabolites [7, 11, 12, 15, 42]. In this work we are interested in overproduction of a secondary metabolite, i.e. the surfactin biosynthesized by NRPS. In a previous work, the problem of lipopeptide synthetase expression was avoided by replacing the surfactin operon promoter by a constitutive one, and this revealed only a small enhancement of surfactin production suggesting that the precursor supply is the problem for the surfactin overproduction in *B. subtilis* 168 derivative strains and not the synthetase expression [3]. To overcome this limitation, we have developed here for the first time an original and fully integrated approach of synthetic biology to overproduce the secondary metabolite surfactin and its precursors.

4.1 Modeling language and knockout prediction for surfactin overproduction

The integrated approach was focused on a modeling language for reaction networks with partial information and on a new knockout prediction algorithm based on abstract interpretation that applies to all models written in this language. We have applied this new knockout prediction algorithm to the overproduction of intracellular leucine production, which in turn was conjectured to increase the production of surfactin in different mutants of *B. subtilis* 168. To achieve our result, we have completely modeled one of the most complex and regulated metabolic pathway of *B. subtilis*, i.e. the branched chain amino acid metabolic pathway, in the new modeling language. When applied to the regulation subnetwork of the *ilv-leu* operon, our prediction algorithm returned four solutions corresponding to four single knockout predictions with respect to the larger model. And when applied to the complete model, 12 plausible single knockout predictions were returned, which included the four previous predictions. From these 12 predictions, we carried-out the six with most direct effect and verified them in vivo. Since, the peptide moiety of surfactin contains four leucine molecules; we have measured the surfactin concentration as an indicator to detect the intracellular level of leucine.

4.2 Metabolic flux prediction and effect on surfactin production

The regulation of the *ilv-leu* operon is dependent on the availability and nature of carbon and nitrogen sources in the chemical medium (the context), since these are involved in various global regulators i.e. CcpA (carbon), TnrA (nitrogen) and CodY (amino acids and Gtp) and other regulators which depends upon the influence of these global regulators and change in concentration of various intracellular metabolites [21, 23–25, 32]. Therefore, our results also depend on the medium that is chosen. To analyze the specific surfactin production the mutants strains were cultivated in two different media, which contain a different composition in carbon and nitrogen sources. Indeed, TSS medium with 16 amino acid is low concentrated in glucose (5 g/L), but very rich in organic nitrogen (4.5 g/L), unlike the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ is very rich in glucose (20 g/L) and mainly contains inorganic nitrogen (3.6 g/L) and a small amount of organic nitrogen (1 g/L yeast extract). As shown in Fig. 4 the differences in medium composition leads to differences in specific surfactin production. The production is increased for all the mutants in glucose rich medium (Landy), and this is also true for the control strain. This finding confirms the prediction made without knockout (penalties 4; shown in Table 1B) where an increase in pyruvate input lead to an increase of leucine production. Pyruvate is a direct result of glycolysis, we

could consider that an increase in glucose concentration in the medium may result in an increase of pyruvate intracellular concentration. Otherwise, the difference in nitrogen supply reveals a very different behavior of the mutants.

4.2.1 Genetic engineering of *B. subtilis* and growth analysis

Most of the regulatory genes involved in the branched chain amino acid metabolic pathway being pleiotropic regulators in *B. subtilis*, so the effect on the growth of the strains in which these genes were deleted has also been studied. Growth analysis of the different mutant strains shows that the deletions of *tnrA* and *Tbox* have no significant impact on the specific growth rate of the strains. These data suggest that no significant influence was contributed by these regulators on the growth of the mutants regardless of the carbon and nitrogen composition of the medium. The disruption of high affinity CodY binding site has no effect on the growth of the strain but the *codY* deletion has detrimental effect on the growth in comparison to BBG111 in TSS medium with 16 amino acids. The delay in the growth of BBG254 in this medium may be due to the impact of CodY on the growth of the strain under carbon limited conditions. Regarding the genes involved in the branched chain amino acid degradation *bcd* and *bkL*, which code for branched chain amino acid dehydrogenase and E3 lipoamide dehydrogenase (*lpdV*) [39] the effects shown are different. Thus, a significant reduction in the growth rate of the strain deleted for *bcd* (BBG253) was observed in comparison to the control strain BBG111 in the TSS medium with 16 amino acids. This has also occurred to a lesser extent in the Landy medium. *bcd* deletion may lead a disturbance in the synthesis of branched-chain fatty acids, which plays an important role in maintaining the fluidity of the membrane lipids [43]. In contrast, in absence of *lpdV*, the production of a α -keto acids occurs, which suggests that *lpdV* is not essential as two other homologs of lipoamide dehydrogenase exist in *B. subtilis* complementing its function [39]. Although, for some of these genes knocked out an effect on growth is observed, all the obtained mutants are viable because these genes are not listed as essential for the cell [44].

4.3 Knockout of the *ilv-leu* operon regulation and effect on surfactin production

The expression of *ilv-leu* operon is regulated by CodY in the presence of branched chain amino acid and is known to respond to high content of GTP present in exponentially growing phase [34]. In the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, there was an increase of seven times in the specific surfactin production with BBG254 (*codY* deleted; r_{16}), which reached an exceptional value of 1.3 g of surfactin produced per gram of dried biomass. This pro-

vides an evidence for more expression of *ilv-leu* in the mutant strains. Thus results in an increase of intracellular pool of leucine. The production of the strain BBG251 (BS_{CodY} deleted; r_1) was not significantly different in this condition than the control strain. But, in TSS medium with 16 amino acids, the difference between CodY mutants and the control strain is more important. An enhancement of 21 times in the specific surfactin production was shown when *codY* is totally deleted (BBG254) and five times when the CodY high affinity binding site is removed (BS_{CodY} ; BBG251), in comparison to BBG111. These results highlight the strong negative regulation provided by CodY on the branched chain amino acids metabolism and especially when the medium is rich in amino acids. Shivers and Sonenshein [21] have reported that the presence of Ile and Val increase the affinity of CodY for its binding site. In the case where CodY remains active, we suppose that it could be able to repress the *ilv-leu* operon via its others binding sites. Indeed, the $P_{\text{ilv-leu}}$ promoter contains four binding sites for CodY, in the promoter-proximal CodY-I+II (nt -84/-32) site that is a high-affinity binding site of CodY, while the promoter-distal CodY-III (nt-154/-107) and CodY-IV (185/-168) sites are low-affinity ones [21]. This shows that modification of CodY regulation in the cell led to an enhancement of specific surfactin production by increasing the production of leucine via the overexpression of *ilv-leu* operon. This observation has been perfectly predicted by our modeling of the $P_{\text{ilv-leu}}$ regulation (r_1 and decreasing the input of CodY, see Table 1A and also in our modeling of the complete metabolic pathway [r_{15}]). Moreover the comparison between the specific surfactin production of the strain BBG254 and the strain BBG251 are consistent regarding the results previously published by Brinsmade et al. [24]. In this study, the authors show that the copies of *ilv-leu* transcript for CodY null mutant strain was 80 times more as well as for high-affinity CodY binding site it was 28 times higher in TSS medium with 16 amino acids. The $P_{\text{ilv-leu}}$ is also subject to a negative regulation by TnrA through its binding to the specific sites. This stimulates DNA binding inhibiting transcription initiation and repressing the *ilv-leu* expression [32]. To avoid this regulation, the prediction suggests deleting the input of TnrA (r_{16} or to suppress its binding on the $P_{\text{ilv-leu}}$; r_7). This repression was disrupted by the deletion of *tnrA* in the strain BBG256. In TSS medium with 16 amino acids, the specific surfactin production was increased twice, but this result was not considered as significant reported to that of the control strain. But in Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, where the concentration of organic nitrogen was less than in the latter medium, a four times increase in specific surfactin production was observed. These results are quite logical since the impact of this regulator is bestowed only in organic nitrogen limited conditions [45].

Intracellular concentration of leucine also downregulates the *ilv-leu* operon using a feedback loop through an attenuator [24, 31]. This 80-bp leucine-dependent stem-loop structure (Tbox) has been identified by the predictions as a target to be deleted (r_{40}). The effect of its deletion on specific surfactin production of the BBG252 strain was clearly better in TSS medium with 16 amino acids (nine-fold increasing) than in the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ (1.5-fold increasing) in comparison to BBG111. Presence of leucine in the TSS medium probably has the effect of inhibiting the expression of *ilv-leu* operon in BBG111, regulation which does not occur in the Landy medium containing only $(\text{NH}_4)_2\text{SO}_4$ as nitrogen source. Brinsmade et al. [24] also found that the copies of *ilv-leu* transcript were 19 times higher in Tbox mutant (SRB87) in comparison to the wild type strain in TSS medium with 16 amino acids.

4.4 Knockout of the genes involved in the leucine degradation

Most of the predictions newly obtained from the complete leucine network were related to the genes involved in the leucine degradation: *ywaA* via r_{47} , *bkL* via r_{11} and *bcd* via r_{10} , and the mechanism which activates the expression of these two latter genes, i.e. expression of BkdR via Ile and Val activation via r_{12} and r_{13} and its binding on the promoter via r_3 or the constitutive expression of the promoter by r_4 . All these knockout predictions have the same effect of interrupting the two stages of reactions that allow the degradation of branched chain amino acids to make fatty acid precursors (Fig. 3). To test all these predictions, *bcd* and *bkL* were deleted in the strains BBG253 and BBG255, respectively. Here, we have chosen not to interrupt *ywaA* because it is also involved in leucine production through the reaction r_{45} . However, YwaA action in this step might be able to be offset by YbgE as these two enzymes are exchangeable and have the same function [35]. In TSS medium with 16 amino acids, an increment of six times in the specific surfactin production was observed with BBG253 strain, but no significant increase was detected with BBG255 strain. This small change observed is perhaps due to the presence of its homolog in *B. subtilis*. Nevertheless, in Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, the effect of knockout on *bcd* and *bkL* led to the same result for both strains with an enhancement of three times in specific surfactin production. Finally, *lpdV* deletion seems to be the right strategy because few effects on growth were observed and the impact on the surfactin production was good.

5 Concluding remarks

All these results discussed above confirm the efficiency and the relevance of the modeling language and the

abstract interpretation algorithm developed for gene knockout prediction. Indeed regarding the medium conditions all the predicted knockouts led to an increase in surfactin production. There are many questions for future work when it comes to the knockout prediction algorithm. First, they could be extended to the whole metabolism of *B. subtilis*, but before this, it would be very helpful to fully automatize the interpretation of the solutions of the constraint solver. This includes better tools that can automatize the presentation of the results. Second, we would like to point out that the quality of the approximation obtained by abstract interpretation depends on the strength of the simplification of the steady state equations, before applying the abstract interpretation. In order to obtain reasonable guarantees for the quality of these simplifications, the algorithm should be investigated in greater depth. The question is whether it can be formulated as a confluent term rewrite system. Third and most importantly, the present prediction algorithm is purely qualitative, so that it cannot predict anything quantitative on the size of the changes. Finding more accurate quantitative abstract interpretations would be an important improvement. New in vivo metabolic engineering experiments should be performed with the different mutants to determine the missing values of model parameters. This in-depth metabolic study will be implemented in the models to make it consistent. In summary, this work opens numerous perspectives, firstly to improve the models and secondly to apply this original approach to more complex secondary metabolites biosynthesis or more complex metabolic pathways such as may be found in eukaryotic organisms.

*This research received funding from the European Union, Marie Curie ITN AMBER, 317338 and partly funded by the University of Lille 1 through the BQR project Biologie synthétique pour la synthèse dirigée de peptides microbiens bioactifs. The authors thank Dr Max Béchet from the ProBioGEM team, Dr Vladimir Bidnenko and Dr Sandrine Auger from INRA Jouy-en-Josas for their contributions in the mutants construction. We are also grateful to the REALCAT platform of the University Lille 1 that allowed the mutants screening. The REALCAT platform is benefiting from a Governmental subvention administrated by the French National Research Agency (ANR) with the contractual reference ANR-11-EQPX-0037. We thank Pr Sonenshein from the Tufts University of Boston for donations of the *B. subtilis* SRB87 and SRB94 strains. And most importantly, we thank our new master student Thibault Etienne for helping us with the time consuming experiments with the constraint solver.*

The authors declare no financial or commercial conflict of interest.

6 References

- [1] Jacques, P., Surfactin and other lipopeptides from *Bacillus* spp., in: *Biosurfactants*, Springer Berlin Heidelberg, 2011, pp 57–91.
- [2] Jauregi, P., Coutte, F., Catiau, L., Lecouturier, D. et al., Micelle size characterization of lipopeptides produced by *B. subtilis* and their recovery by the two-step ultrafiltration process. *Sep. Purif. Technol.* 2013, 104, 175–182.
- [3] Coutte, F., Leclère, V., Béchet, M., Guez, J.-S. et al., Effect of pps disruption and constitutive expression of *sfA* on surfactin productivity, spreading and antagonistic properties of *Bacillus subtilis* 168 derivatives. *J. Appl. Microbiol.* 2010, 109, 480–491.
- [4] Béchet, M., Castéra-Guy, J., Guez, J.-S., Chihib, N.-E. et al., Production of a novel mixture of mycosubtilins by mutants of *Bacillus subtilis*. *Bioresour. Technol.* 2013, 145, 264–270.
- [5] Sun, H., Bie, X., Lu, F., Lu, Y. et al., Enhancement of surfactin production of *Bacillus subtilis* fmbR by replacement of the native promoter with the *Pspac* promoter. *Canad. J. Microbiol.* 2009, 55, 1003–1006.
- [6] Liu, J.-F., Yang, J., Yang, S.-Z., Ye, R.-O. et al., Effects of different amino acids in culture media on surfactin variants produced by *Bacillus subtilis* TD7. *Appl. Biochem. Biotech.* 2012, 166, 2091–2100.
- [7] Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z. N. et al., Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* 2004, 427, 839–843.
- [8] Covert, M. W., Palsson, B. O., Constraints-based models: Regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.* 2003, 221, 309–325.
- [9] Florez, L. A., Gunka, K., Polanfa, R., Tholen, S. et al., SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. *BMC Syst. Biol.* 2011, 5, 5.
- [10] Price, N. D., Reed, J. L., Palsson, B. O., Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2004, 2, 886–897.
- [11] Kim, J., Reed, J. L., OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* 2010, 4, 53.
- [12] Ranganathan, S., Suthers, P. F., Maranas, C. D., OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* 2010, 6, e1000744.
- [13] Burgard, A. P., Pharkya, P., Maranas, C. D., OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 2003, 84, 647–657.
- [14] Wiechert, W., ¹³C metabolic flux analysis. *Metab. Eng.* 2001, 3, 195–206.
- [15] Goelzer, A., Brikci, F. B., Martin-Verstraete, I., Noirot, P. et al., Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst. Biol.* 2008, 2, 20.
- [16] Papin, J. A., Stelling, J., Price, N. D., Klamt, S. et al., Comparison of network-based pathway analysis methods. *Trends Biotechnol.* 2004, 22, 400–405.
- [17] Gruchattka, E., Hadicke, O., Klamt, S., Schütz, V. et al., In silico profiling of *Escherichia coli* and *Saccharomyces cerevisiae* as terpenoid factories. *Microb. Cell Fact.* 2013, 12, 1–18.
- [18] Jungreuthmayer, C., Zanghellini, J., Designing optimal cell factories: Integer programming couples elementary mode analysis with regulation. *BMC Syst. Biol.* 2012, 6, 103.
- [19] John, M., Nebut, M., Niehren, J., Knockout prediction for reaction networks with partial kinetic information, in: *Verification, Model Checking, and Abstract Interpretation*, Springer 2013, pp. 355–374.
- [20] Niehren, J., John, M., Versari, C., Coutte, F. et al., Qualitative reasoning about reaction networks with partial kinetic information. *Computational Methods for Systems Biology*, Nantes, France, 2015, hal-01163391, 12.
- [21] Shivers, R. P., Sonenshein, A. L., Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol. Microbiol.* 2004, 53, 599–611.
- [22] Tojo, S., Satomura, T., Morisaki, K., Deutscher, J. et al., Elaborate transcription regulation of the *Bacillus subtilis* *ilv-leu* operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA. *Mol. Microbiol.* 2005, 56, 1560–1573.
- [23] Villapakkam, A. C., Handke, L. D., Belitsky, B. R., Levdikov, V. M. et al., Genetic and biochemical analysis of the interaction of *Bacillus subtilis* CodY with branched-chain amino acids. *J. Bacteriol.* 2009, 191, 6865–6876.
- [24] Brinsmade, S. R., Kleijn, R. J., Sauer, U., Sonenshein, A. L., Regulation of CodY activity through modulation of intracellular branched-chain amino acid pools. *J. Bacteriol.* 2010, 192, 6357–6368.
- [25] Brinsmade, S. R., Sonenshein, A. L., Dissecting complex metabolic integration provides direct genetic evidence for CodY activation by guanine nucleotides. *J. Bacteriol.* 2011, 193, 5637–5648.
- [26] Mader, U., Schmeisky, A. G., Florez, L. A., Stülke, J., SubtiWiki – a comprehensive community resource for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* 2012, 40, 1278–1287.
- [27] Landy, M., Warren, G., Rosenman, M. S., Colio, L., Bacillomycin an antibiotic from *Bacillus subtilis* active against pathogenic fungi. *Exp. Biol. Med.* 1948, 67, 539–541.
- [28] Tanaka, K., Henry, C. S., Zinner, J. F., Jolivet, E. et al., Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res.* 2012, 41, 1–13.
- [29] Nicolas, P., Mader, U., Dervyn, E., Rochat, T. et al., Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* 2012, 335, 1103–1106.
- [30] Sambrook, J. R., Russel, D. D. W., in: *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press 2001.
- [31] Grandoni, J. A., Zahler, S. A., Calvo, J. M., Transcriptional regulation of the *ilv-leu* operon of *Bacillus subtilis*. *J. Bacteriol.* 1992, 174, 3212–3219.
- [32] Fujita, Y., Satomura, T., Tojo, S., Hirooka, K., CcpA-mediated catabolite activation of the *Bacillus subtilis* *ilv-leu* operon and its negation by either CodY- or TnrA-mediated negative regulation. *J. Bacteriol.* 2014, 196, 3793–3806.
- [33] Belitsky, B. R., Sonenshein, A. L., Contributions of multiple binding sites and effector-independent binding to CodY-mediated regulation in *Bacillus subtilis*. *J. Bacteriol.* 2011, 193, 473–484.
- [34] Ratnayake-Lecamwasam, M., Serror, P., Wong, K.-W., Sonenshein, A. L., *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. *Gene Dev.* 2001, 15, 1093–1103.
- [35] Thomaidis, H. B., Davison, E. J., Burston, L., Johnson, H. et al., Essential bacterial functions encoded by gene pairs. *J. Bacteriol.* 2007, 189, 591–602.
- [36] Molle, V., Nakaura, Y., Shivers, R. P., Yamaguchi, H. et al., Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J. Bacteriol.* 2003, 185, 1911–1922.
- [37] Mader, U., Hennig, S., Hecker, M., Homuth, G., Transcriptional organization and posttranscriptional regulation of the *Bacillus subtilis* branched-chain amino acid biosynthesis genes. *J. Bacteriol.* 2004, 186, 2240–2252.
- [38] Hatfield, G. W., Umbarger, H. E., Threonine deaminase from *Bacillus subtilis*. II. The steady state kinetic properties. *J. Biol. Chem.* 1970, 245, 1742–1747.
- [39] Debarbouille, M., Gardan, R., Arnaud, M., Rapoport, G., Role of BkdR, a transcriptional activator of the SigL-dependent isoleucine and

- valine degradation pathway in *Bacillus subtilis*. *J. Bacteriol.* 1999, 181, 2059–2066.
- [40] Duitman, E. H. *Nonribosomal peptide synthesis in Bacillus subtilis*. Ph.D. thesis, University Library Groningen 2003.
- [41] Otero, J. M., Nielsen, J., Industrial Systems Biology. *Biotechnol. Bioeng.* 2010, 105, 439–460.
- [42] Sohn, S. B., Kim, T. Y., Park, J. M., Lee, S. Y., In silico genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnol. J.* 2010, 5, 739–750.
- [43] Kaneda, T., Iso- and anteiso-fatty acids in bacteria: Biosynthesis, function, and taxonomic significance. *Microbiol. Rev.* 1991, 55, 288–302.
- [44] Commichau, F. M., Pietack, N., Stülke, J., Essential genes in *Bacillus subtilis*: A re-evaluation after ten years. *Mol. Biosyst.* 2013, 9, 1068–1075.
- [45] Fisher, S. H., Regulation of nitrogen metabolism in *Bacillus subtilis*: Vive la difference! *Mol. Microbiol.* 1999, 32, 223–232.