

# **Investigating the metabolism of non-*aureus* staphylococci relevant to prosthetic joint infection**

**Teresa Díaz Calvo**

Norwich Medical School

University of East Anglia

This dissertation is submitted for the degree  
of Doctor of Philosophy

2020



**In collaboration with Oxford Brookes University**

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

# Abstract

Prosthetic joint infection (PJI) is a complication of joint replacement that occurs when bacteria adhere to the surface of a prosthetic joint and form a biofilm (Becker *et al.* 2014; Tande *et al.* 2014). Treatment is expensive and aggressive, constituting a major burden to the healthcare system and to the patients (Becker *et al.* 2014; Tande *et al.* 2014). Non-*aureus* staphylococci account for approximately 30% of all cases (Becker *et al.* 2014), with *S. epidermidis* being the main species involved (Becker *et al.* 2014), showing an increasing pathogenic potential (Uribe-Alvarez *et al.* 2016). Despite this, little is known about the mechanism of infection, which is assumed to be biofilm formation. This work describes how the first highly curated genome-scale metabolic model for *S. epidermidis* RP62A was constructed, manually curated, validated against experimental data and analysed with linear-programming techniques to explore the metabolism of cells living in joints. We defined routes for production of energy, planktonic biomass and biofilm polymers during growth on nutrients found in synovial fluid and under the range of conditions encountered across the biofilm structure: the results obtained indicated that the metabolic network re-arranges itself, varying the uptake and metabolism of glucose and amino acids in response to environmental changes and highlighted the importance of the uptake and catabolism of citrulline for ATP production, a pathway that, to our knowledge, has not been described before in this context. This work also provided an explanation for experimental observations where a decrease in the production of biofilm was observed *in vitro* upon glutamate deprivation, linking its catabolism with the synthesis of ATP, and suggested that the cell's ability to modify the level of de-acetylated residues in biofilm exopolysaccharides is an important feature of biofilm formation. Finally, it exemplified how metabolic modelling can be useful in anticipating regulatory patterns leading to optimal bacterial growth strategies in different environments. This work is being developed further with a focus on informing the pathogenesis of non-*aureus* staphylococci in PJI.

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. John Wain and Dr. Gemma Langridge, who played a key part in my decision to pursue this PhD. I am enormously grateful to them for sharing their knowledge and their passion for science as well as for encouraging me to explore my own ideas while providing invaluable support. They have been truly inspirational and have contributed greatly to my professional and personal development. I am also thankful to all present and former members of the MMRL team at the Norwich Medical School and the members of our research group at the Quadram Institute for creating such a nice working environment and making coming to work every day a really enjoyable experience.

I would like to thank Dr. Mark Poolman for his great patience while teaching me metabolic modelling from scratch. I have been truly lucky to count on such a highly-skilled and knowledgeable mentor. I am very grateful for his input to this project, our great, inspirational discussions and our close on-going collaboration. I would also like to thank Prof. David Fell and the rest of the members of the Cell System Modelling Group at Oxford Brookes University for being so welcoming and making my stays at Oxford always prolific and very enjoyable and for introducing me to Nepalese food.

I would like to thank Dr. Noemi Hernandez Tejera for her incredible hard work and support in obtaining the experimental data needed to validate this work. I cannot thank her enough for this, for her friendship and for brightening up those cloudy winter days.

I am also very thankful to Dr. Dipali Singh for her encouragement and for providing crucial practical support, her constant patience and continuous smile, as well as for being such a great friend. This last year would have been very different without her.

A big thank-you goes to Dr. Enriqueta Garcia Gutierrez for our stimulating discussions in the pub and for involving me in her research on *S. epidermidis*, as well as for her quick and practical master class on thesis-formatting, which made this writing up process less stressful and more enjoyable.

I would like to take this chance to acknowledge the contribution of Prof. Iain McNamara to the clinical side of this project and to thank him for his on-going support.

I would also like to thank my partner, Oliver P. Lawer, for his support during this journey and for providing crucial distraction (and delicious food) when it was most needed.

I would like to dedicate this thesis to my family, whom I miss every day and who always have been, are, and will be with me on every step of the way.

This project was partly funded by The University of East Anglia and The Orthopaedics Charitable Trust Fund.

# Contents

1	General introduction.....	1
1.1	Project background and motivation.....	1
1.2	Aims and structure .....	1
1.3	Non- <i>aureus</i> staphylococci.....	2
1.3.1	Staphylococci and their classification.....	2
1.3.2	Genetic diversity of NAS .....	3
1.3.3	NAS species and association with disease: current perspective .....	6
1.4	Staphylococcal metabolism .....	6
1.4.1	The electron transport chain .....	7
1.4.2	Metabolic states in staphylococci .....	8
1.4.2.1	First metabolic state.....	8
1.4.2.2	Second metabolic state.....	9
1.4.3	Metabolism in absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> .....	9
1.4.4	Biofilm metabolism: production and utilisation of acetoin and butanediol.....	10
1.4.5	Metabolic features of other clinically relevant phenotypes: Small Colony Variants..	10
1.5	Minimal growth requirements for staphylococci .....	10
1.5.1	Previously published minimal growth requirements for <i>S. epidermidis</i> RP62A .....	11
1.6	Nitrogen assimilation and amino acid catabolism in staphylococci.....	11
1.6.1	Nitrogen assimilation .....	11
1.6.2	The effect of glucose on amino acid catabolism.....	12
1.6.3	Differential amino acid utilisation on planktonic cultures and biofilms .....	13
1.7	Biofilm production in staphylococci .....	14
1.7.1	Molecular mechanisms of biofilm formation .....	15
1.7.2	Environmental conditions encountered by bacteria growing in joints .....	16

1.7.2.1	Composition of the synovial fluid.....	16
1.7.3	Composition of biofilms growing in joints.....	17
1.7.4	Structure and synthesis of PIA in <i>S. epidermidis</i> .....	18
1.7.5	Metabolic features of staphylococcal biofilms.....	20
1.8	Prosthetic Joint Infection.....	20
1.8.1	The scope of the problem.....	21
1.8.2	Etiological causes.....	22
1.8.3	Diagnosis.....	23
1.8.3.1	Typing of NAS.....	23
1.8.3.2	Clinical diagnosis: the decision-making process.....	24
1.8.4	Treatment.....	24
1.9	Metabolic reconstructions of staphylococci.....	25
2	General methods.....	27
2.1	Introduction to mathematical modelling of metabolism.....	27
2.2	Structural metabolic modelling.....	28
2.2.1	The steady state assumption.....	29
2.2.2	Null space analysis.....	30
2.2.3	Elementary modes analysis.....	31
2.2.4	Linear programming-based analysis.....	32
2.3	Software and metabolic modelling tools.....	34
2.3.1	The Python programming language.....	34
2.3.2	ScrumPy: metabolic modelling with Python.....	34
2.4	Construction, curation and analysis of the genome-scale model.....	34
2.4.1	Genome annotation, databases and tools.....	35
2.4.2	Model construction and structure.....	36

2.4.2.1	'Top-level' module .....	37
2.4.2.2	Automatically generated reactions .....	37
2.4.2.3	Transporters .....	37
2.4.2.4	Electron transport chain .....	38
2.4.2.5	Additional reactions.....	38
2.4.3	Model curation.....	39
2.4.3.1	Correction of atomically unbalanced reactions .....	40
2.4.3.2	Material consistency of the model.....	41
2.4.3.3	Synthesis of polymers.....	41
2.4.3.4	Addition of transporters for cell biomass components .....	42
2.4.4	Fundamental validation of the model.....	43
2.4.4.1	Model analysis for fundamental validation.....	43
2.4.4.2	Model-wide conservation of mass.....	44
2.4.4.3	Model-wide conservation of energy and redox .....	44
2.4.4.4	Feasibility of biomass production.....	44
2.4.4.5	Definition of auxotrophies and essential media components .....	45
2.4.4.6	Other quality checks .....	45
2.5	General properties of the model.....	45
3	Fundamental characterisation of the genome-scale model .....	48
3.1	Introduction .....	48
3.2	Methods.....	48
3.2.1	Model analysis for ATP production.....	48
3.3	Results.....	49
3.3.1	Characterisation of the electron transport chain .....	49

3.3.2	Metabolic responses for ATP production under a range of environmental conditions	52
3.3.2.1	ATP production from glucose	52
3.3.2.2	ATP production from glutamate	61
3.3.2.3	ATP production from acetate	64
3.3.3	Biofilm energy metabolism: production of acetoin and butanediol and their utilisation for ATP synthesis	66
3.3.3.1	Production of acetoin and butanediol	66
3.3.3.2	Utilisation of metabolic by-products for ATP production	67
3.3.4	Metabolic responses for production of planktonic biomass under a range of environmental conditions	68
3.4	Discussion	70
3.4.1	Characterisation of the ETC	70
3.4.2	Metabolic responses for ATP under a range of environmental conditions	70
3.4.2.1	ATP production from glucose	70
3.4.2.2	ATP production from glutamate	71
3.4.2.3	ATP production from acetate	72
3.4.3	Biofilm energy metabolism: production of acetoin and butanediol and utilisation for ATP synthesis	72
3.4.3.1	Production of acetoin and butanediol	72
3.4.3.2	Utilisation of metabolic by-products for ATP production	73
3.4.4	Metabolic responses for production of planktonic biomass under a range of environmental conditions	73
3.5	Conclusion	74
4	Model refinement and validation: minimal growth requirements	75
4.1	Introduction	75
4.1.1	Use of experimental data to validate GSMs of staphylococci	75

4.1.2	Comparison between minimal growth requirements for <i>S. epidermidis</i> RP62A defined <i>in vitro</i> and by LP-based analysis of the GSM.....	77
4.2	Methods.....	79
4.2.1	Experimental assessment of minimal growth requirements in <i>S. epidermidis</i> RP62A	79
4.2.1.1	Minimal media composition.....	79
4.2.1.2	Inoculum and bacterial strains.....	81
4.2.1.3	Experimental setup.....	81
4.2.2	Assessment of the impact of amino acid deprivation on biofilm formation.....	82
4.2.2.1	Crystal violet biofilm staining method.....	82
4.2.3	Comparison between experimental data and <i>in silico</i> results.....	82
4.2.3.1	Model analysis for essentiality of media components.....	82
4.2.3.2	Study of the biosynthetic potential encoded in the genome of the organism.....	83
4.2.3.3	Identification of possible mutations in the laboratory strain.....	83
4.2.3.4	Model refinement.....	84
4.2.3.5	Computation of the effect of single amino acid deprivation on biomass production by LP-based analysis.....	84
4.3	Results.....	85
4.3.1	Experimental assessment of requirements for vitamins and amino acids in <i>S. epidermidis</i> RP62A.....	85
4.3.2	Impact of amino acid deprivation on biofilm formation.....	88
4.3.3	Comparison between experimental data and <i>in silico</i> results.....	89
4.3.3.1	Model validation and refinement.....	89
4.3.3.2	Model analysis for interpretation of experimental results.....	91
4.4	Discussion.....	93
4.4.1	Experimental assessment of requirements for amino acids in <i>S. epidermidis</i> RP62A.....	93

4.4.2	Impact of amino acid deprivation on biofilm formation.....	94
4.4.3	Comparison between experimental data and <i>in silico</i> results .....	95
4.5	Conclusion .....	98
5	Analysis-guided experimental validation of the GSM on nitrogen metabolism and amino acid utilisation for biomass production.....	99
5.1	Introduction.....	99
5.1.1	Investigation of proline essentiality.....	99
5.1.1.1	Proline biosynthesis .....	99
5.1.1.2	The osmoprotective role of proline.....	100
5.1.1.3	Impact of carbon catabolite repression on proline biosynthesis .....	100
5.1.2	Use of LP-based analysis to inform experimental design for hypothesis testing.....	101
5.2	Methods .....	101
5.2.1	Model analysis for the study of the <i>in silico</i> utilisation of amino acids for biomass production .....	101
5.2.1.1	Overall uptake and excretion of amino acids and their uptake to demand ratio for biomass production.....	101
5.2.1.2	Contribution of individual amino acids to the total nitrogen and carbon uptake for biomass production.....	102
5.2.1.3	Growth on single amino acids as nitrogen sources .....	102
5.2.1.4	Nitrogen assimilation: glutamate biosynthesis from $\text{NH}_4^+$ .....	102
5.2.2	Experimental design based on the LP-based analysis results .....	103
5.2.2.1	Bacterial inoculum.....	103
5.2.2.2	Media composition.....	103
5.2.2.3	Experimental setup.....	105
5.3	Results .....	105
5.3.1	<i>In silico</i> utilisation of amino acids for biomass production .....	105

5.3.1.1	Overall uptake and excretion of amino acids and their uptake to demand ratio for biomass production .....	106
5.3.1.2	Contribution of individual amino acids to the total nitrogen and carbon uptake for biomass production .....	107
5.3.1.3	Growth on single amino acids as nitrogen sources.....	111
5.3.1.4	Nitrogen assimilation: glutamate biosynthesis from $\text{NH}_4^+$ .....	113
5.3.2	Experimental work for model validation on amino acid utilisation for biomass production .....	114
5.3.2.1	Assessment of growth in the absence of proline .....	114
5.3.2.2	Assessment of growth on single amino acids (alanine, arginine and glutamate) and on a mixture of them.....	114
5.3.2.3	Assessment of growth on a mixture of amino acids defined as non-suitable nitrogen sources for RP62A.....	115
5.3.2.4	Assessment of growth on $\text{NH}_4^+$ as sole nitrogen source.....	116
5.3.2.5	Experimental results .....	116
5.4	Discussion.....	123
5.4.1	<i>In silico</i> utilisation of amino acids for biomass production.....	123
5.4.1.1	Overall uptake and excretion of amino acids and their uptake to demand ratio for biomass production .....	123
5.4.1.2	Contribution of individual amino acids to the total nitrogen and carbon uptake	124
5.4.1.3	Growth on single amino acids as nitrogen sources.....	125
5.4.1.4	Nitrogen assimilation: glutamate biosynthesis from $\text{NH}_4^+$ .....	125
5.4.2	Experimental work for model validation on amino acid utilisation for biomass production .....	126
5.4.2.1	Assessment of growth in the absence of proline .....	126
5.4.2.2	Assessment of growth on single amino acids (alanine, arginine and glutamate) and on a mixture of them.....	127

5.4.2.3	Assessment of growth on a mixture of amino acids defined as non-suitable nitrogen sources for RP62A and on $\text{NH}_4^+$ as sole nitrogen source .....	128
5.4.2.4	The effect of Glc on amino acid metabolism.....	128
5.5	Conclusion .....	129
6	Applying model analysis to the study of the metabolism of RP62A cells growing on prosthetic joints.....	131
6.1	Introduction.....	131
6.1.1	Synthesis of PIA in the RP62A GSM.....	132
6.1.2	Composition of the <i>in silico</i> synovial fluid.....	133
6.2	Methods .....	134
6.2.1	Model analysis for production of ATP, planktonic biomass and PIA in joints.....	134
6.2.1.1	Metabolic responses for ATP and biomass production in the presence and absence of $\text{O}_2$ and $\text{NO}_3^-$ .....	134
6.2.1.2	Simulated variation in the $\text{O}_2$ concentration during ATP and biomass production	134
6.2.1.3	Generation of submodels and computation of elementary modes .....	135
6.3	Results .....	136
6.3.1	Production of ATP to cover the GAM and NGAM energy demand .....	136
6.3.1.1	Metabolic responses for ATP production in synovial fluid in the presence and absence of $\text{O}_2$ and $\text{NO}_3^-$ .....	136
6.3.1.2	Simulated variation in the $\text{O}_2$ concentration during ATP production: responsive reactions, subnetworks and elementary modes.....	143
6.3.2	Production of PIA.....	146
6.3.2.1	Metabolic responses for production of PIA in synovial fluid in the presence and absence of $\text{O}_2$ and $\text{NO}_3^-$ .....	146
6.3.2.2	Simulated variation in the $\text{O}_2$ concentration during production of PIA: responsive reactions, subnetworks and elementary modes.....	153
6.3.3	Production of planktonic biomass .....	157

6.3.3.1	Metabolic responses for production of planktonic biomass in synovial fluid in the presence and absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> .....	158
6.3.3.2	Simulated variation in the O <sub>2</sub> concentration during production of planktonic biomass: responsive reactions, subnetworks and elementary modes .....	160
6.4	Discussion.....	162
6.4.1	ATP production in joints .....	162
6.4.1.1	Metabolic responses for ATP production in synovial fluid in the presence and absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> .....	162
6.4.1.2	Simulated variation in the O <sub>2</sub> concentration during ATP production: responsive reactions, subnetworks and elementary modes .....	162
6.4.2	PIA production in joints .....	163
6.4.2.1	Metabolic responses for production of PIA in synovial fluid in the presence and absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> .....	163
6.4.2.2	Simulated variation in the O <sub>2</sub> concentration during production of PIA: responsive reactions, subnetworks and elementary modes .....	164
6.4.3	Production of planktonic biomass in joints .....	166
6.4.3.1	Metabolic responses for production of planktonic biomass in synovial fluid in the presence and absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> .....	166
6.4.3.2	Simulated variation in the O <sub>2</sub> concentration during production of planktonic biomass: responsive reactions, subnetworks and elementary modes .....	166
6.5	Conclusion.....	167
7	General discussion.....	169
7.1	Comparison of GSMs of staphylococci.....	169
7.2	Results overview.....	170
7.3	Future work.....	174
7.4	Conclusions .....	176
8	Bibliography .....	178
9	Appendices.....	192

9.1	Appendix A: additional material .....	192
9.1.1	Python modules for model construction .....	192
9.1.1.1	Model construction .....	192
9.1.1.2	Deleted reactions and metabolites.....	192
9.1.1.3	Corrected reactions and metabolites .....	193
9.1.2	Python modules for constructive analysis of the model .....	193
9.1.2.1	General properties of the model.....	193
9.1.2.2	General validation of the model.....	193
9.1.2.3	ETC stand-alone module .....	193
9.1.2.4	Feasibility of production of single biomass components .....	193
9.1.2.5	Feasibility of biomass production.....	194
9.1.2.6	ATP production .....	194
9.1.2.7	Essentiality of media components.....	194
9.1.3	Python modules for functional analysis of the model .....	194
9.1.3.1	Essentiality of amino acids and effect of amino acid deprivation on biomass production. ....	194
9.1.3.2	Utilisation for biomass production and assimilation of inorganic N.....	195
9.1.3.3	Oxygen scan .....	195
9.2	Appendix B: biomass composition.....	195
9.3	Appendix C: experimental data.....	198
9.3.1	Minimal growth requirements.....	198
9.3.2	Impact of amino acid deprivation on biofilm formation.....	201
9.4	Appendix D: model refinement.....	202
9.4.1	Comparison between experimental data and <i>in silico</i> results on the essentiality of media components.....	202
9.4.1.1	Biosynthesis of amino acids .....	202

9.4.1.2	Biosynthesis of vitamins .....	206
---------	--------------------------------	-----

# List of Figures

Figure 1-1 Reconstruction of staphylococcal phylogenetic trees based on the 16S rRNA and agrB genes .....	4
Figure 1-2 Circular layout of the phylogeny generated from the analysis of 16S rRNA protein gene sequences from 225 isolates of NAS, Norwich Medical School, UEA.....	5
Figure 1-3 Schematic representation of the glucose catabolism of <i>S. aureus</i> during the exponential and post-exponential growth phases.....	9
Figure 1-4 Schematic representation of the steps involved in the process of biofilm formation.....	16
Figure 1-5 Schematic representation of the biosynthetic process of the biofilm matrix polymer PIA in staphylococci .....	19
Figure 1-6 Schematic showing a total hip arthroplasty in place with relevant structures highlighted .....	21
Figure 1-7 Knee revision at Norfolk & Norwich University Hospital showing actions and decision points .....	25
Figure 2-1 Schematic representation of a simple set of reactions .....	28
Figure 2-2 Elementary modes identified in the metabolic network described in Figure 2-1 .....	32
Figure 2-3 Modular structure of the <i>S. epidermidis</i> RP62A GSM.....	37
Figure 2-4 Diagram describing the steps that integrate the process of metabolic model construction and refinement. ....	40
Figure 3-1 Reactions of the ETC representing the menaquinone-dependent $\text{NO}_3^-$ reduction mechanism described for staphylococci. ....	50
Figure 3-2 Reactions of the ETC alternative mechanism of $\text{NO}_3^-$ reduction via cytochromes described for staphylococci.....	51
Figure 3-3 ATP production from Glc in the presence of $\text{O}_2$ when total flux through the system was minimised. ....	54
Figure 3-4 ATP production from Glc in the presence of $\text{NO}_3^-$ when total flux through the system was minimised. ....	55
Figure 3-5 ATP production from Glc in the presence of $\text{O}_2$ and $\text{NO}_3^-$ when the import of Glc was minimised. ....	56
Figure 3-6 ATP production in the presence of Glc and the absence of $\text{O}_2$ and $\text{NO}_3^-$ for both objective functions: minimisation of net total flux through and minimisation of Glc consumption) .....	57
Figure 3-7 ATP production in the presence of Glc and the absence of $\text{O}_2$ and $\text{NO}_3^-$ when total flux through the system was minimised and flux through the reaction catalysed by the short chain acyl-CoA-dh (EC 1.3.8.1) was blocked. ....	60
Figure 3-8 ATP production from Glt in the presence of $\text{O}_2$ when total flux through the system was minimised. ....	62
Figure 3-9 ATP production from Glt in the presence of $\text{O}_2$ and $\text{NO}_3^-$ when the import of Glt was minimised. ....	63

Figure 3-10 ATP production in the presence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> with Ac as sole C source and minimising total flux through the system.....	65
Figure 4-1 Growth curves for <i>S. epidermidis</i> RP62A in four representative modified MM <sup>-</sup> medium samples presenting no apparent delay or a slight delay in growth in comparison to cultures on standard MM medium.....	86
Figure 4-2 Growth curves for <i>S. epidermidis</i> RP62A in all modified MM <sup>-</sup> medium samples exhibiting an apparent growth delay in comparison to cultures on standard MM medium.....	87
Figure 4-3 Relative levels of biofilm formation in <i>S. epidermidis</i> RP62A cultures growing on standard MM medium and modified MM <sup>-</sup> medium lacking single amino acids.....	88
Figure 5-1 Growth curves for <i>S. epidermidis</i> RP62A cultures in modified MM medium samples in comparison to cultures in standard MM medium over a period of 7 days.....	117
Figure 5-2 Growth curves for <i>S. epidermidis</i> RP62A cultures in modified MM medium samples in comparison to cultures in standard MM medium over a period of 7 days.....	118
Figure 6-1 ATP production in synovial fluid in the presence of O <sub>2</sub> when total flux through the system was minimised.....	137
Figure 6-2 ATP production in synovial fluid in the presence of NO <sub>3</sub> <sup>-</sup> when total flux through the system was minimised.....	138
Figure 6-3 ATP production in synovial fluid in the absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> when total flux through the system was minimised.....	139
Figure 6-4 ATP production in synovial fluid in the absence of electron acceptors under depletion of citrulline when total flux through the system was minimised.....	140
Figure 6-5 Change in flux for responsive reactions involved in aerobic and anaerobic respiration during ATP production in synovial fluid over variation in the availability of O <sub>2</sub> when NO <sub>3</sub> <sup>-</sup> was available and total net flux through the system was minimised.....	144
Figure 6-6 Production of PIA in synovial fluid in the presence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> when total flux through the system was minimised.....	147
Figure 6-7 Production of PIA in synovial fluid in the presence of O <sub>2</sub> and absence of NO <sub>3</sub> <sup>-</sup> when total flux through the system was minimised.....	148
Figure 6-8 Production of PIA in synovial fluid in the presence of NO <sub>3</sub> <sup>-</sup> and absence of O <sub>2</sub> when total flux through the system was minimised.....	149
Figure 6-9 Production of PIA in synovial fluid in the absence of electron acceptor when total flux through the system was minimised.....	150
Figure 6-10 Change in flux for responsive importers of medium components involved in PIA synthesis in synovial fluid over variation in the availability of O <sub>2</sub> in the absence of NO <sub>3</sub> <sup>-</sup> when total net flux through the network was minimised.....	155
Figure 9-1 Growth curves for <i>S. epidermidis</i> RP62A in all modified MM <sup>-</sup> medium samples presenting no apparent delay in growth in comparison to cultures in standard MM medium (purple).....	198
Figure 9-2 Growth curves for <i>S. epidermidis</i> RP62A in modified MM <sup>-</sup> medium samples without single vitamins compared to cultures in standard MM medium (purple).....	199

# List of Tables

Table 1-1 Common etiological causes of PJI .....	23
Table 2-1 General properties of the GSM for <i>S. epidermidis</i> RP62A.....	46
Table 3-1 Model responses for production of ATP from Glc considering two alternative analysis objectives in the presence and absence of electron acceptors.....	53
Table 3-2 Model response for production of ATP from Glc considering two alternative analysis objectives in the absence of electron acceptors.....	59
Table 3-3 P/O ratios calculated for the LP analysis solutions obtained for ATP production from Glc considering two alternative analysis objectives in the presence of O <sub>2</sub> .....	61
Table 3-4 Characterisation of the model behaviour for production of planktonic cell biomass in the presence and absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> considering minimisation of total net flux through the network as the objective of the analysis.....	69
Table 3-5 Characterisation of the model behaviour for production of planktonic cell biomass in the presence and absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> upon blocking flux through the short chain acyl-CoA-dh reaction while minimising total net flux through the network.....	69
Table 4-1 Comparison of minimal growth requirements for <i>S. aureus</i> N315 identified by analysis of four different GSMs .....	76
Table 4-2 Comparison between the biosynthetic potential for amino acids encoded in the genome of <i>S. epidermidis</i> RP62A, experimentally reported auxotrophies and auxotrophies defined by model analysis. ....	78
Table 4-3 Composition of the standard MM medium formulated for minimal growth requirement experiments.....	80
Table 4-4 Summary of the <i>in vitro</i> growth effects observed upon removal of single amino acids from the MM medium compared with the experimental observations produced by Hussain <i>et al.</i> (1991). .....	87
Table 4-5 Summary of the effects caused by removal of single amino acids from the standard MM medium on <i>S. epidermidis</i> RP62A cell growth and cell phenotype. ....	89
Table 4-6 Comparison between the biosynthetic potential for amino acids encoded in the genome of <i>S. epidermidis</i> RP62A, amino acids auxotrophies reported experimentally and proposed explanations for the discrepancies encountered. ....	90
Table 4-7 Summary of the effect caused by removal of single amino acids on the objective value and Glc consumption of the LP-based analysis solutions for biomass production in MM medium when the objective of the analysis was to minimise the total net flux through the network.....	91
Table 4-8 <i>In silico</i> effect of the removal of amino acids and NH <sub>4</sub> <sup>+</sup> on the objective value and the Glc uptake rate of responses for production of cell biomass in standard MM medium and modified MM medium when the objective of the analysis was to minimise the total net flux through the network. .....	92

Table 5-1 Concentration of amino acids present in tMM medium and their respective concentration in standard MM medium.....	104
Table 5-2 Overall uptake or excretion of $\text{NH}_4^+$ and amino acids during in silico production of planktonic biomass in standard MM medium and their uptake to demand ratios in the presence of $\text{O}_2$ . .....	106
Table 5-3 Contribution of each amino acid to the total N taken up for synthesis of planktonic biomass when $\text{NH}_4^+$ and all 19 amino acids were present in the standard MM medium in the presence of $\text{O}_2$ and total net flux through the network was minimised. ....	108
Table 5-4 Uptake and excretion of N for the production of planktonic biomass when amino acids were either absent (i) or present (ii) in MM medium in the presence of $\text{O}_2$ while total net flux through the network was minimised.....	109
Table 5-5 Contribution of each amino acid to the total C uptake for synthesis of planktonic biomass when $\text{NH}_4^+$ and all 19 amino acids were present in the standard MM medium in the presence of $\text{O}_2$ when total net flux through the network was minimised. ....	110
Table 5-6 Uptake and excretion of C for the production of planktonic biomass when amino acids were either absent (i) or present (ii) in MM medium in the presence of $\text{O}_2$ when total net flux through the network was minimised.....	111
Table 5-7 Variation on the total net flux of the LP solution when just a single N source was available at a time for production of planktonic biomass in the presence of $\text{O}_2$ when total net flux through the network was minimised.....	112
Table 5-8 Main features of the LP-based analysis solutions obtained when Ala, Arg and Glt were available as sole N sources for biomass production in MM medium in the presence of $\text{O}_2$ when total net flux through the network was minimised. ....	115
Table 5-9 Main features of the LP-based analysis solution obtained when a mixture of Ala, Arg and Glt was available as sole N source for biomass production in MM medium in the presence of $\text{O}_2$ when total net flux through the network was minimised. ....	115
Table 5-10 $A_{600}$ values for <i>S. epidermidis</i> RP62A cultures in modified MM medium containing different N sources after 7 days of incubation.....	119
Table 5-11 $A_{600}$ values for <i>S. epidermidis</i> RP62A cultures in MM medium and MM <sup>-</sup> medium without Pro after 2 and 7 days of incubation .....	119
Table 5-12 Summary of the in silico effect of amino acid and $\text{NH}_4^+$ removal from the standard MM medium on the total net flux through the system and the Glc demand for production of planktonic biomass in the presence of $\text{O}_2$ .....	122
Table 5-13 Comparison of results derived from LP-based analysis of the model and the experimental work performed during this study.....	130
Table 6-1 Characterisation of the model behaviour for production of ATP in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors. .....	142

Table 6-2 Main N and C sources utilised and by-products excreted during production of ATP in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.....	142
Table 6-3 Characterisation of the model behaviour for production of ATP in the absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> with Glc and NH <sub>4</sub> <sup>+</sup> as sole available C and N sources when total net flux through the network was minimised. ....	143
Table 6-4 Summary of data obtained with O <sub>2</sub> scans for ATP production in synovial fluid in the presence and absence of NO <sub>3</sub> <sup>-</sup> when total net flux through the network was minimised.....	143
Table 6-5 Characterisation of the model behaviour for production of PIA in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors..	152
Table 6-6 Main N and C sources utilised and by-products excreted during production of PIA in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.....	152
Table 6-7 Characterisation of the model behaviour for production of PIA in the absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> with Glc and NH <sub>4</sub> <sup>+</sup> as sole C and N sources when total net flux through the network was minimised. ....	153
Table 6-8 Summary of data obtained in the O <sub>2</sub> scans for PIA production in synovial fluid when total net flux through the network was minimised in the presence and absence of NO <sub>3</sub> <sup>-</sup> .....	154
Table 6-9 Characterisation of the model behaviour for production of planktonic biomass in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors. ....	158
Table 6-10 Main N and C sources utilised and by-products excreted during production of planktonic biomass in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.....	159
Table 6-11 Characterisation of the model behaviour for production of planktonic biomass in the absence of O <sub>2</sub> and NO <sub>3</sub> <sup>-</sup> with Glc and NH <sub>4</sub> <sup>+</sup> as sole C and N sources when total net flux through the network was minimised.....	160
Table 6-12 Summary of data obtained with O <sub>2</sub> scans for production of planktonic biomass in synovial fluid and the presence and absence of NO <sub>3</sub> <sup>-</sup> when total net flux through the network was minimised. ....	161
Table 7-1 General properties of curated GSMs available for staphylococci .....	169
Table 9-1 Biomass composition for <i>S. epidermidis</i> derived from data corresponding to <i>S. aureus</i> and modified on basis to biochemical data currently available for <i>S. epidermidis</i> . ....	196
Table 9-2 Comparison between the biosynthetic potential for vitamins encoded in the genome of <i>S. epidermidis</i> RP62A, vitamin auxotrophies reported experimentally and their essentiality according to the GSM after curation .....	200
Table 9-3 Relative levels of biofilm formation detected for RP62A cultures growing in different media cultures growing in different media and variations observed with respect to cultures in the standard MM medium. ....	201

# URLs

ACT	<a href="https://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act">https://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act</a>
BioCyc	<a href="http://biocyc.org/">http://biocyc.org/</a>
BRENDA	<a href="http://www.brenda-enzymes.org/">http://www.brenda-enzymes.org/</a>
eQuilibrator	<a href="http://equilibrator.weizmann.ac.il/">http://equilibrator.weizmann.ac.il/</a>
GLPK	<a href="http://www.gnu.org/software/glpk/">http://www.gnu.org/software/glpk/</a>
Gnuplot	<a href="http://www.gnuplot.info/">http://www.gnuplot.info/</a>
KEGG	<a href="https://www.genome.jp/kegg/">https://www.genome.jp/kegg/</a>
NumPy	<a href="http://www.numpy.org/">http://www.numpy.org/</a>
Prokka	<a href="https://github.com/tseemann/prokka">https://github.com/tseemann/prokka</a>
Python	<a href="http://docs.python.org/">http://docs.python.org/</a>
Roary	<a href="https://github.com/sanger-pathogens/Roary">https://github.com/sanger-pathogens/Roary</a>
SeaView	<a href="http://doua.prabi.fr/software/seaview">http://doua.prabi.fr/software/seaview</a>
SciPy	<a href="http://www.scipy.org/">http://www.scipy.org/</a>
ScrumPy	<a href="http://mudshark.brookes.ac.uk/ScrumPy">http://mudshark.brookes.ac.uk/ScrumPy</a>

# Abbreviations and formulae

<i>Elements and metabolites</i>	<i>Common name</i>
Ac	Acetate
AcCoA	Acetyl-CoA
Acetald	Acetaldehyde
Aceto-AcCoA	Acetoacetyl-CoA
Acetyl-P	Acetyl-phosphate
ADP	Adenosine-diphosphate
AIP	Autoinducing peptide
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartate
ATP	Adenosine-triphosphate
BPGA	1,3-diphosphoglycerate
C	Carbon
Carbamoyl-P	Carbamoyl phosphate
Cit	Citrate
CisAco	Cis-aconitate
CO <sub>2</sub>	Carbon dioxide
Cys	Cysteine
DHAP	Dihydroxyacetone phosphate
EtHO	Ethanol
FBP	Fructose bisphosphate
Form	Formate
Fum	Fumarate
F6P	Fructose-6-phosphate
GAP	Glyceraldehyde 3-phosphate
Glc	Glucose

GlcN-F6P	Glucosamine-fructose-6P
GlcN-1P	Glucosamine-1P
GlcN-6P	Glucosamine-6P
Glt	Glutamate
Glutamyl-P	Glutamyl phosphate
Gly	Glycine
G6P	Glucose-6-phosphate
His	Histidine
HMG-CoA	3-hydroxy-3-methyl-glutaryl-CoA
Homo-Cys	Homo-cysteine
Homo-Ser	Homo-serine
H <sub>2</sub> O	Water
Ile	Isoleucine
IsoCit	Isocitrate
Lac	Lactate
Leu	Leucine
Lys	Lysine
Mal	Malate
Mena	Menaquinones
Met	Methionine
MSCRAMMs	Microbial surface components recognizing adhesive matrix molecules
N	Nitrogen
NAcGal	<i>N</i> -acetyl-D-galactosamine
NAcGlc	<i>N</i> -acetylglucosamine
NAcGlc-1P	<i>N</i> -acetylglucosamine-1-phosphate
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NH <sub>3</sub>	Ammonia
NH <sub>4</sub> <sup>+</sup>	Ammonium

NO <sub>2</sub> <sup>-</sup>	Nitrite
NO <sub>3</sub> <sup>-</sup>	Nitrate
OAA	Oxaloacetate
Ox-Cyto	Cytochromes-C-oxidized
O <sub>2</sub>	Oxygen
Phe	Phenylalanine
Pi	Phosphate
PPi	Diphosphate
Pro	Proline
Pyr	Pyruvate
Red-Cyto	Cytochromes-C-reduced
Red-Mena	Reduced menaquinones
PEP	Phosphoenol pyruvate
PGA	3-phospho-D-glycerate
PIA	Polysaccharide intercellular adhesine
PIA1	Polysaccharide intercellular adhesine initial polymer
PIA2	Polysaccharide intercellular adhesine de-acetylated
PNAG	Poly- <i>N</i> -acetylglucosamine
Pyr	Pyruvate
Ser	Serine
SucCoA	Succinyl-S-CoA
Suc	Succinate
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
UDP-NAcGlc	Uridine-diphosphate- <i>N</i> -acetylglucosamine-1-phosphate
UDP	Uridine-diphosphate
UTP	Uridine-triphosphate
Val	Valine
2PG	2-phospho-D-glycerate

2KG	2-oxoglutarate
<b><i>Enzymatic reactions</i></b>	<b><i>Common name</i></b>
Ac kinase	Acetate kinase
Ac-CoA ligase	Acetate-CoA ligase
Acetald-dh	Acetaldehyde dehydrogenase
Acyl-CoA-dh	Acyl-CoA dehydrogenase
Ala-aminotransferase	Alanine aminotransferase
Ala-dh	Alanine dehydrogenase
Cytochrome-c NO <sub>2</sub> <sup>-</sup> reductase	Cytochromes-C nitrite reductase
GlcN-F6P aminotransferase	Glucosamine-fructose-6P aminotransferase
Glt-dh	Glutamate dehydrogenase
GOGAT	Glutamate synthase
GS	Glutamine synthetase
Isocitrate-dh	Isocitrate dehydrogenase
P-acetyl transferase	Phosphate-acetyl transferase
Pta-AckA pathway	Phosphate-acetyl transferase-acetate kinase pathway
Pyr-dh	Pyruvate dehydrogenase
Pyr Form-lyase	Pyruvate formate lyase
Quinol-NO <sub>3</sub> <sup>-</sup> reductase	Quinol-nitrate reductase
Suc-CoA-synthase	Succinyl-S-CoA synthase
Thr-aldolase	Threonine aldolase
2-hydroxyacyl-CoA-dh	2-hydroxyacyl-CoA dehydrogenase
2KG-dh	2-oxoglutarate dehydrogenase
<b><i>Others</i></b>	<b><i>Common name</i></b>
CFU	Colony forming unit
EM	Elementary mode
ETC	Electron transport chain
GAM	Growth-associated maintenance cost
GSM	Genome-scale metabolic model
LP	Linear programming

MALDI-TOF MS	Matrix assisted laser desorption/ionization time-of flight mass spectrometry
MLST	Multi-locus sequence typing
NAS	Non- <i>aureus</i> staphylococci
NGAM	Non growth-associated maintenance cost
PJI	Prosthetic joint infection
PMF	Proton motive force
P/O	Phosphate/oxygen ratio
PPP	Pentose phosphate pathway
RC	Robust clusters
SCVs	Small colony variants
STs	Sequence types
TCA	Tricarboxylic acid cycle



# 1 General introduction

## 1.1 Project background and motivation

Non-*aureus* staphylococci (NAS) are a diverse group of organisms which are poorly understood, yet important in disease and, specifically, very important in PJI. This project was born as part of the core research in NAS taking place at the Medical Microbiology Research Laboratory of the Norwich Medical School in 2016. By then, public and commercially funded research was being undertaken with the aim of commercially exploiting the study of genotypic and phenotypic diversity in this group of organisms. During these studies it became evident that the understanding of the mechanisms underlying the ability of NAS to colonise prosthetic joints was lacking. Thus this project aims to move forward our understanding of the pathogenicity of NAS through the generation of a curated model of the well described *S. epidermidis* strain RP62A.

The expression of metabolic genes leads to synthesis of the enzymes catalysing chemical reactions occurring in the bacterial cell, with the exception of spontaneous reactions, which occur without enzymatic catalysis. Translating the bacterial genome into reactions allows us to define the metabolic network associated with the bacterial phenotype. Since it is by modifying their metabolism that bacteria adapt to new environments, understanding the metabolic features of NAS became one of the main interests of the research group. Analysis of structural metabolic models, which are mathematical representations of metabolic networks provides in-depth insight into these complex systems and ultimately helps to understand the metabolic basis of cellular phenotypes. These models also allow for the integration of phenotypic data sets in order to improve accuracy. Thus building and analysing a model for the *S. epidermidis* strain RP62A was a challenging but exciting opportunity to expand this research, as well as to establish collaborations with experts in the field and to test the worth of applying these techniques to a wider range of NAS strains and to other research projects.

## 1.2 Aims and structure

Specifically, three objectives were identified and pursued from the beginning of the project:

- a) To construct, curate and validate a high quality genome-scale metabolic model for *S. epidermidis* RP62A, a genome sequenced representative of the main NAS species involved in PJI.
- b) To analyse the model in order to define the metabolic pathways that allow this organism to grow in joints.
- c) To identify the metabolic mechanisms involved in biofilm formation in *S. epidermidis* RP62A.

This thesis is organised in the following manner:

**Chapter 1** presents the project background and gives a comprehensive introduction to NAS and PJI.

**Chapter 2** describes the basis of the mathematical modelling methodology used in the construction, curation and analysis of a genome-scale model of *S. epidermidis*, as well as the main aspects of these processes and the general properties of the model after its fundamental validation.

**Chapter 3** describes how the model is analysed in order to reproduce and investigate physiological features and metabolic strategies of the organism.

**Chapter 4** describes how the system is analysed for defining minimal growth requirements and the results obtained are compared with experimental data and used for further validation and refinement of the model.

**Chapter 5** describes how model analysis is used as a guiding-tool to define experiments for further validation of the system on the metabolism of nitrogen and amino acids, as well as the comparison of the *in silico* and *in vitro* results.

**Chapter 6** describes how the model is analysed for production of energy, planktonic biomass and biofilm polymers under the conditions encountered by cells living in the intra-articular space of joints.

**Chapter 7** is a general discussion of the results presented in this thesis and how this relates to the ability of *S. epidermidis* to cause infection.

## **1.3 Non-aureus staphylococci**

### **1.3.1 Staphylococci and their classification**

Staphylococcal bacterial cells are non-motile and spherical, with an approximate diameter of 1.0  $\mu\text{m}$  and usually appear as grape-like clusters. Staphylococci are Gram-positive microorganisms and most of them are catalase positive. They are facultative anaerobes that typically grow at temperatures between 18°C and 40°C and are able to tolerate high salt concentrations (Somerville 2016). The genus *Staphylococcus* is composed of multiple species that have been traditionally divided between ‘coagulase-positive’ and ‘coagulase-negative’ staphylococci. However, this division corresponded to the clinical need to differentiate between pathogenic and non-pathogenic organisms and is not based on phylogenetic relationships (Becker *et al.* 2014; Somerville 2016).

Historically, staphylococci were first divided into two groups, in 1884, by their ability to produce pigment: *S. aureus* (golden colonies) and *S. albus* (white colonies) (Cowan *et al.* 1954; Becker *et al.* 2014). Soon after that, the need to differentiate between pathogenic and non-pathogenic

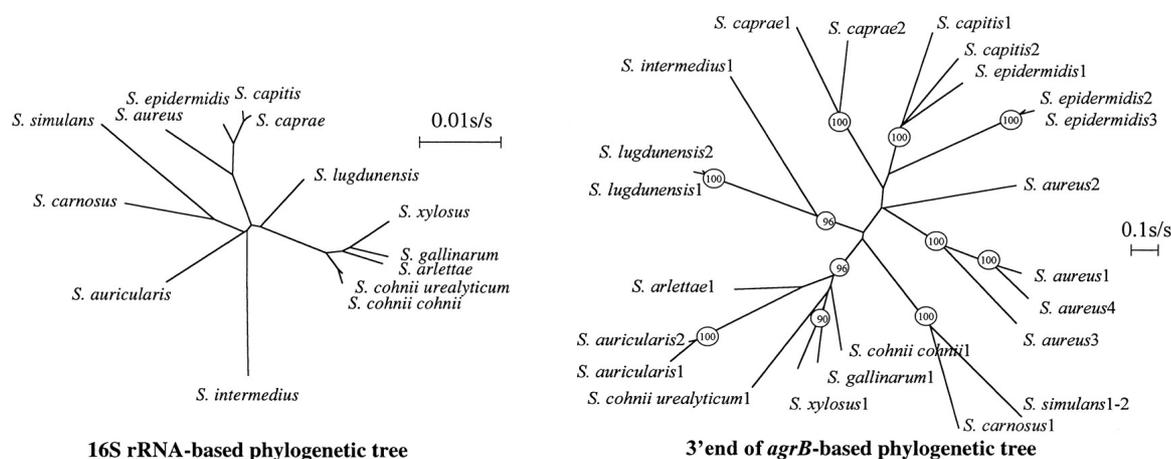
staphylococci led to the development of other typing protocols. In 1940, production of coagulase was considered a reliable method for distinguishing between pathogenic (coagulase-positive) and non-pathogenic (coagulase-negative) staphylococcal species (Fairbrother 1940). It is worth noting that *S. aureus* was the only known coagulase-positive species at the time (Becker *et al.* 2014). Since then, other coagulase-positive staphylococcal species have been described, and some *S. aureus* strains have been shown to be coagulase-negative (Somerville 2016). Although *S. aureus* is still considered as the most important pathogenic species, the association between several “coagulase-negative staphylococci” and disease has been established and is reported to be increasing (Pereira 1962; Becker *et al.* 2014). However, despite its less than perfect association with infection, this classification scheme has prevailed. The genus *Staphylococcus* is currently composed of 47 species (Becker *et al.* 2014; Somerville 2016) and *S. aureus* is the most notorious and the best known of all. *S. aureus* are highly virulent organisms and are known to be involved in several pathologies, including skin and respiratory infections, food poisoning and bacteraemia, which can lead to a life threatening conditions (Jarraud *et al.* 2002). Other species traditionally classified as coagulase-positive are: *S. simiae*, *S. intermedius*, *S. delphini*, *S. lutrae* and *S. pseudointermedius*, while *S. scheiferi* comprises coagulase-positive and coagulase-negative subspecies and the presence of coagulase in *S. agnetis* has been described as variable (Becker *et al.* 2014). Despite being significantly different, the rest of staphylococcal species have traditionally been grouped together as coagulase-negative staphylococci.

Due to the reasons mentioned above, it seems no longer reasonable or practical to continue referring to all non-*aureus* staphylococcal species as ‘coagulase-negative staphylococci’. Therefore, the term used for them in this thesis is simply ‘non-*aureus* staphylococci’ or NAS.

### 1.3.2 Genetic diversity of NAS

Classification of staphylococci into species is most commonly based on the comparison of genetic sequences within the 16S rRNA operon, which are conserved throughout prokaryotes. This method compares the unknown sequence of a given organism to sequences of type strains defined by classical methods (Mincheol Kim 2014). Although this approach generates a reproducible phylogeny it does not always match phylogenetic trees based on other widespread housekeeping genes. For example, the *agr* locus (Robinson *et al.* 2005; Ikuo *et al.* 2014) has also been used for the phylogenetic classification of staphylococci (Dufour *et al.* 2002; Robinson *et al.* 2005): this is a quorum sensing accessory gene regulator consisting on the genes *agrB*, *agrD*, *agrC*, and *agrA*, and the regulatory effector molecule RNAlII, and is involved in the regulation of many virulence factors in staphylococci, being widespread (Tan *et al.* 2018) and well conserved (Robinson *et al.* 2005; Ikuo *et al.* 2014) in these species. When a typing study compared sequences related to the *agr* regulon in several *Staphylococcus* species (Dufour *et al.* 2002) and constructed phylogenetic trees from the different *agr* alleles, the interspecies relationships based on the *agrB* and *agrC* genes seemed to have

high discriminatory power and yield reliable phylogenetic groups. However, subsequent comparison of these phylogenies with a tree based on the 16S rRNA gene showed that some isolates defined as *S. epidermidis* by the 16S method appeared to be closer to *S. capitis* by *agrB* (Figure 1-1).

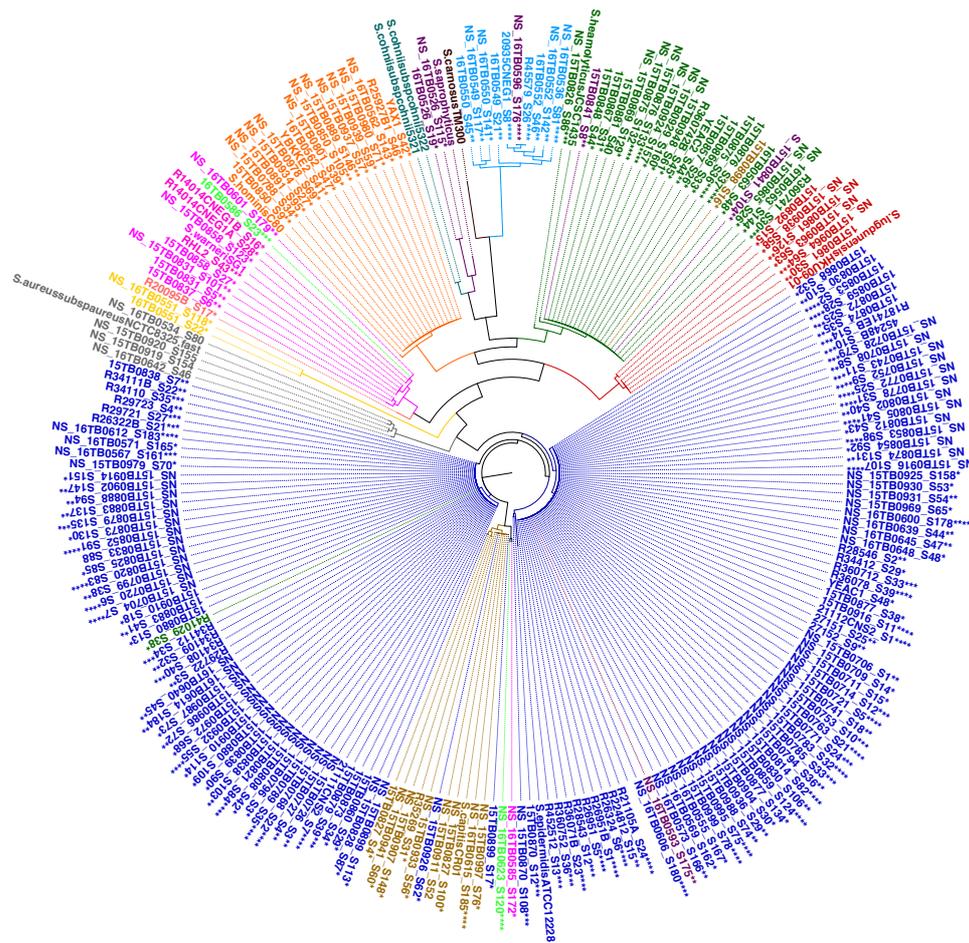


**Figure 1-1 Reconstruction of staphylococcal phylogenetic trees based on the 16S rRNA and *agrB* genes**

Data from Philippe Dufour *et al.* 2002 and reproduced here with permission.

Furthermore, different regions of the genome appear to give different phylogenies suggesting horizontal gene exchange. The situation is further complicated because the different NAS species present variable levels of genetic diversity (Miragaia *et al.* 2007; Bückle *et al.* 2017), with the population structure of some species being far more diverse than others. Two good examples are the two species *S. epidermidis* and *S. carnosus* (Miragaia *et al.* 2007; Bückle *et al.* 2017): in 2007, a study typed 217 *S. epidermidis* nosocomial isolates from several countries (Miragaia *et al.* 2007) using multi-locus sequence typing (MLST) and described 74 sequence types (STs) within 1 major and 8 minor clonal complexes, with recombination estimated to contribute to clonal diversity approximately twice as much as point mutations. When a similar approach was taken to study population diversity in *S. carnosus* from various sources (Bückle *et al.* 2017), 44 isolates were divided into only 9 STs but the population had a marked clonal structure and, in this case, genetic diversity was mainly attributed to point mutations.

Recent, unpublished, studies performed at the Norwich Medical School, UEA (commercial confidentiality applies) question the current definition of NAS species. Using whole genome sequencing, a phylogeny was generated from the 16S ribosomal protein genes (described by (Hug *et al.* 2016)) for 225 NAS isolates collected to represent diversity. The groups generated, or “robust clusters” (RCs), were compared to the current species identification by mass spectrometry analysis (MALDI-TOF MS) (Figure 1-2).



**Figure 1-2 Circular layout of the phylogeny generated from the analysis of 16S rRNA protein gene sequences from 225 isolates of NAS, Norwich Medical School, UEA**

The isolates were coloured according to their species characterization by MALDI-TOF MS in the following manner: brown = *S. capitis*; dark blue = *S. epidermidis*; light green = *S. cohnii*; dark green = *S. haemolyticus*; grey = *S. aureus*; orange = *S. hominis*; pink = *S. warneri*; purple = *S. carnosus*; red = *S. ludgdunensis*. Those isolates in yellow and light blue did not match any specific species profile according to the MALDI-TOF MS database. The number of asterisks shows next to each isolate's name indicates its capacity to form a biofilm under the conditions tested: \*(weak/non biofilm former); \*\*(moderate biofilm former); \*\*\* (strong biofilm former); and \*\*\*\* (very strong biofilm former). This figure is reproduced here with permission from the study team and is under commercial confidentiality.

These data (Figure 1-2) show the wide genetic diversity found within the NAS group, which is particularly striking for the *S. epidermidis* isolates (dark blue), which seem to fall within at least two well-differentiated clusters. Again, isolates classified as *S. capitis* (light brown) appear to be genetically closer to *S. epidermidis* than to any other NAS species, while isolates characterized as *S. simulans* (light blue) seem to fall far away from any other group. Notice how some clusters contain isolates of different colours, hence, belonging to different species according to the current species definition of NAS.

The RCs do not match precisely species as defined by mass spectrometry but, for urinary tract infection do match, more strongly than species, the clinical outcome (data not shown). When tested for biofilm formation there were some RCs which were unable to form biofilms but most clusters contained both strong and weak biofilm forming isolates. This suggests that biofilm formation is not directly inherited, and so is not solely related to gene content, but more likely is subjected to regulation of metabolism.

It is becoming clearer and clearer that our current lack of understanding of the biology of different species within the NAS group is because the species themselves are poorly defined, making it difficult to establish associations between NAS species and disease. Studying the metabolic diversity of NAS will advance our understanding of their biology and so ultimately improve clinical management of patients infected by them.

### **1.3.3 NAS species and association with disease: current perspective**

NAS comprise a diverse range of staphylococcal species, the majority of which are ubiquitous commensals of the skin and the mucous membranes but can behave as opportunistic pathogens under favourable circumstances, representing one of the major nosocomial pathogens (Becker *et al.* 2014). Furthermore, some NAS are currently known to be displaying increasing antimicrobial resistance, which might lead to a rise in their virulence and association with infection (Aggarwal *et al.* 2014; Morgenstern *et al.* 2016). They are particularly associated with the use of indwelling medical devices, which are frequently used in modern medicine (Becker *et al.* 2014): *S. epidermidis* is the main NAS species associated with PJI, as well as with other foreign body-related infections, endocarditis and neonatal infections (Becker *et al.* 2014; Argemi *et al.* 2015; Xue *et al.* 2015). *S. haemolyticus* has been reported in cases of foreign body-related infections (including PJI) and neonatal infections (Becker *et al.* 2014; Argemi *et al.* 2015). Both, *S. epidermidis* and *S. haemolyticus* are the two species more often associated with infections but also more frequently reported as contaminants in clinical microbiology laboratories (Argemi *et al.* 2015). *S. hominis*, *S. capitis*, *S. warneri* (Arciola *et al.* 2005; Von Eiff *et al.* 2006; Campoccia *et al.* 2010; Arciola *et al.* 2012) and *S. caprae* (Allignet *et al.* 1999; Tande *et al.* 2014) are also known to be involved in foreign body-related infections and PJI. Finally, *S. lugdunensis* is frequently reported in patients with endocarditis and PJI (Becker *et al.* 2014; Tande *et al.* 2014) while *S. saprophyticus* is involved in urinary tract infections (Kuroda *et al.* 2001; Becker *et al.* 2014).

## **1.4 Staphylococcal metabolism**

As we understand better the grouping of isolates into species then our understanding of staphylococcal metabolism, by species, will shed light over the metabolic diversity of clinically

relevant phenotypes (such as biofilms or Small Colony Variants (SCVs)) and the metabolic basis of antibiotic resistance and niche or host adaptation. For example, data previously obtained with the analysis of metabolic models of staphylococci has already been used to identify new potential drug targets (Becker *et al.* 2005; Heinemann *et al.* 2005) or to develop typing schemes for pathogenic strains of *S. aureus* (Bosi *et al.* 2016). Before starting the process of model construction, it is convenient to review some of the main physiological features of staphylococci and their metabolism.

### 1.4.1 The electron transport chain

The electron transport chain (ETC) of prokaryotes consists of a set of redox reactions transferring electrons from donors to acceptors, coupled with the translocation of protons across the cell membrane. This generates a proton concentration gradient or chemiosmotic gradient and the subsequent proton motive force (PMF) needed to drive phosphorylation of ADP to ATP by the ATP synthase. Many prokaryotes are able to regulate expression of the ETC components in response to changes in the environment, such as variation in the availability of electron acceptors, leading to the utilization of different electron transport pathways (Anraku 1988). Generally,  $O_2$  is the preferred electron acceptor. However, several species are capable of utilizing alternative compounds in its absence, such as  $NO_3^-$ ,  $NO_2^-$ , fumarate or sulphate, although their reduction does not provide as much energy (Unden *et al.* 1997). Staphylococci are able to utilize both  $O_2$  and  $NO_3^-$  as final electron acceptors, which allows for certain versatility in energy production under diverse environmental conditions (Sasarman *et al.* 1974), (Burke *et al.* 1975; Heinemann *et al.* 2005; Uribe-Alvarez *et al.* 2016). The literature on this matter suggests the presence of two independent but complementary mechanisms for  $NO_3^-$  reduction in staphylococci: i) a mechanism functionally linked to the oxidation of menaquinones, which can work without direct involvement of cytochromes (Sasarman *et al.* 1974; Uribe-Alvarez *et al.* 2016) and ii) a mechanism in which  $NO_3^-$  is directly reduced by cytochromes, which get oxidised, and are then reduced again by oxidising menaquinones (Sasarman *et al.* 1974; Burke *et al.* 1975; Heinemann *et al.* 2005). Furthermore, experimental results, published by Uribe-Alvarez *et al.* 2016, have confirmed differential expression of cytochromes and a menaquinone-dependent  $NO_3^-$  reductase in *S. epidermidis* under aerobic, microaerobic and anaerobic conditions (Uribe-Alvarez *et al.* 2016), indicating that cytochromes are highly expressed in the presence of  $O_2$  but this expression is reduced under microaerobic conditions, and is almost absent in anaerobiosis while expression of the  $NO_3^-$  reductase increases in microaerobic conditions, and is high during anaerobiosis. This implies that cytochromes play a small role in the ETC in absence of  $O_2$ , and hence the menaquinone-dependent  $NO_3^-$ -reducing mechanism is favoured *in vitro* under the experimental conditions used for testing.

The ATP/NADH ratio indicates how many ATP molecules are generated per NADH molecule entering the electron transport chain and it is a commonly used indicator of the efficiency of the oxidative phosphorylation in cells. An ATP/NADH ratio of 2 (Wilkinson 1997; Heinemann *et al.*

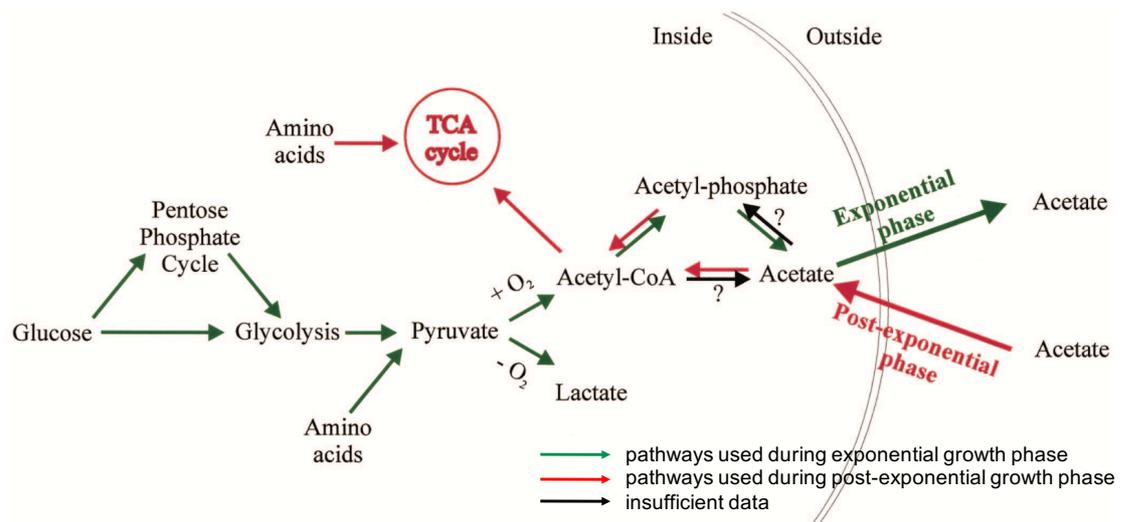
2005) has generally been assumed for other staphylococci like *S. aureus*. Another important bioenergetic parameter is the P/O ratio which describes how much ATP is produced by the movement of an electron pair through the ETC leading to reduction of  $\frac{1}{2}$  O<sub>2</sub> to H<sub>2</sub>O (Garrett 2010); reported estimates vary between 2.5 and 3 (Garrett 2010; Ferrier 2014), with values between this range being typically found when the TCA cycle is active.

## 1.4.2 Metabolic states in staphylococci

Under aerobic conditions, two metabolic states have been experimentally described in staphylococci, specifically in *S. aureus*:

### 1.4.2.1 First metabolic state

Catabolic repression is a common feature of many (but not all) bacteria. It ensures that their metabolism is adapted for quick utilisation of preferred C sources, such as glucose (Glc), by inhibiting the synthesis of catabolic enzymes for other available substrates like amino acids etc (Deutscher 2008). This behaviour has been described in the past by several authors as the ‘glucose effect’ (Blumenthal 1972; Somerville *et al.* 2002; Somerville *et al.* 2003), and was for a long time attributed to wasteful ‘overflow’ metabolism, but is now believed to be linked to “quick growth” producing high growth yields (Kuroda *et al.* 2001; Somerville *et al.* 2003) and cell densities (El-Mansi 2004). ‘Overflow metabolism’ has been observed in fast-growing bacterial, yeast, fungal and mammalian cells and is described as ‘wasteful’ due to the incomplete oxidation of growth substrates even in the presence of O<sub>2</sub>, leading to excretion of Lac, Ac or EtHO. In staphylococci, Glc presence leads to repression of the TCA cycle and Glc being metabolised to Ac, resulting in Ac accumulation in the media. This first metabolic state is observed during exponential growth and involves the incomplete oxidation of Glc to Pyr and further oxidation to Ac via acetyl-P in the Pta-AckA pathway, thus generating ATP by substrate level phosphorylation (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016; Halsey *et al.* 2017) (Figure 1-3). This process is less energy-efficient than performing the TCA cycle and cellular respiration (Vazquez 2018) but an incomplete oxidative strategy could be better suited to support fast growth, since it involves a smaller number of reactions, thus requiring the action of fewer enzymes than full-oxidative metabolism, which not only reduces protein investment but also saves on limited cellular space (Molenaar *et al.* 2009; Basan *et al.* 2015). In 2005, Heinemann *et al.* reproduced this metabolic state in a model of *S. aureus* N315 by artificially constraining the flux through the aconitate hydratase reaction of the TCA cycle (Heinemann *et al.* 2005).



**Figure 1-3 Schematic representation of the glucose catabolism of *S. aureus* during the exponential and post-exponential growth phases**

Figure from Philippe Somerville *et al.* 2003 reproduced here with permission.

### 1.4.2.2 Second metabolic state

When C availability is limited, organisms adapt in order to optimise the ATP yield per C atom by following a more protein-costly strategy which includes glycolysis, the TCA cycle and cellular respiration. This second metabolic state occurs during post-exponential growth, and has been observed in *S. aureus* when the concentration of Glc (or Glt) is reduced (Somerville *et al.* 2003), (Sadykov *et al.* 2013; Somerville 2016). At this point, metabolism transitions between energy production via substrate level phosphorylation to oxidative phosphorylation. In this case, cells adapt to the new conditions by taking up the Ac previously excreted to the media, converting it to AcCoA and feeding it into the TCA cycle, which is now de-repressed and utilised to maximize production of ATP and biosynthetic intermediates (Somerville *et al.* 2003; Somerville 2016) (Figure 1-3). Studies on *S. aureus* strains presenting mutations that cause lack of acetate catabolism have shown no alteration on survival during stationary-phase or production of virulence factors but reported reduced growth yield (Kuroda *et al.* 2001; Somerville *et al.* 2003).

### 1.4.3 Metabolism in absence of $O_2$ and $NO_3^-$

In the absence of electron acceptors, staphylococci are known to perform mixed acid (Lac, Ac, Form) and butanediol fermentation (Fuchs *et al.* 2007), with Glc catabolic end products varying depending on the growth state and the experimental conditions (Sivakanesan *et al.* 1980). The data available regarding the anaerobic metabolism of *S. epidermidis* is very limited: in a study focused on Ser and Glc catabolism where a *S. epidermidis* strain was isolated from a contaminated culture of *Peptococcus prevotii*, the utilisation of Ser as a C source for growth was high under anaerobic conditions in a rich undefined medium but was shown to decrease substantially when the Glc concentration was increased to 0.5% w/v, while the fermentation of Glc was favoured. This strain was specifically

reported to ferment Glc to Lac and trace amounts of Ac, Form and CO<sub>2</sub> (Sivakanesan *et al.* 1980). However, since the origin of the strain used was uncertain and the media utilised through these experiments was undefined, this data should be considered with caution.

#### **1.4.4 Biofilm metabolism: production and utilisation of acetoin and butanediol**

A particular metabolic strategy described in staphylococci seems to be especially important during biofilm formation: cells growing in anaerobic or microaerobic environments, as is the case with staphylococcal biofilms, have been reported to mainly present a fermentative behaviour leading to excretion of Ac, Form and/or Lac (Resch *et al.* 2005; Zhu *et al.* 2007), but at a certain point of the exponential growth phase, cells divert Pyr metabolism to production and excretion of acetoin and butanediol. These metabolites are then available for use to obtain energy during the stationary growth phase, being taken up and converted to AcCoA (Zhu *et al.* 2007). Inhibition of the butanediol pathway or the acetolactate decarboxylase, which reduces acetolactate to acetoin, has been shown to prevent biofilm formation in *S. aureus* (Cassat *et al.* 2006). Since both acetoin and butanediol are neutral compounds, this strategy is believed to help counteract excessive media acidification caused by the excretion of acidic by-products (Yao *et al.* 2005; Xiao *et al.* 2007) while preventing a redox imbalance by regenerating NAD (Zhu *et al.* 2007).

#### **1.4.5 Metabolic features of other clinically relevant phenotypes: Small Colony Variants**

Small Colony Variants (SCVs) are mutant staphylococcal cells that constitute a sub-population frequently isolated in persistent antibiotic resistant infections (Proctor *et al.* 1994; McNamara *et al.* 2000). This phenotype is characterized by presenting approximately one tenth of the wildtype colony size, lack of pigmentation, slow growth, altered carbohydrate utilization patterns and reduced hemolytic activity. These phenotypic changes have been related to an interruption of the ETC (Proctor *et al.* 1994) (McNamara *et al.* 2000). SCVs present an essentially fermentative carbohydrate metabolism, fermenting Pyr to Lac. Using metabolic modelling, Heinemann *et al.* reproduced this behavior *in silico* by artificially removing the heme-requiring cytochrome-b oxidase in a model of *S. aureus* N315, which interrupted the ETC and led to the system fermenting Glc to Lac (Heinemann *et al.* 2005).

### **1.5 Minimal growth requirements for staphylococci**

Throughout the years, researchers have studied minimal growth requirements for staphylococci and these have been used to curate and validate metabolic reconstructions, including the one generated in this project. Previous work demonstrates that certain strains of *S. aureus* can be ‘trained’ to overcome certain, apparent, amino acid auxotrophies by gradually simplifying the culture media,

ultimately leading to growth without amino acids utilising  $\text{NH}_3$  as the sole N source (Gladstone 1937; Knight 1937). Staphylococci are known to have a complex regulatory network which has been shown to inhibit growth in minimal media through pathway repression (Becker *et al.* 2005; Heinemann *et al.* 2005; Lee *et al.* 2009; Bosi *et al.* 2016), a phenomenon that can be reversed after several passages in simplified media (Gladstone 1937; Heinemann *et al.* 2005; Bosi *et al.* 2016). Repression of biosynthetic pathways under certain circumstances (e.g. growth in rich media or inside a host where amino acids are largely available) can lead to varying growth phenotypes (Gladstone 1937; Knight 1937; Heinemann *et al.* 2005; Lee *et al.* 2009; Bosi *et al.* 2016) and, as a result, several authors have reported differential amino acid and vitamin requirements for staphylococcal species, not only between but also within strains (Knight 1937; Emmett *et al.* 1975).

In summary, while an organism's genome might encode for all the enzymes required for the synthesis of a certain compound, if these are not transcribed or translated at sufficient levels at a given time, then that specific compound might be required in the media for growth (conditionally essential) but is not truly essential.

### **1.5.1 Previously published minimal growth requirements for *S. epidermidis* RP62A**

In 1991, Hussain, Hastings and White performed growth requirement experiments on a set of *S. epidermidis* strains which included RP62A (Hussain *et al.* 1991). Using a minimal medium previously described by Gladstone in 1937 (Gladstone 1937) and removing one single amino acid at a time from a mixture of 18 amino acids, bacterial growth was estimated by eye after 18, 24 and 48 hours. Results indicated that by time 18 hours, no growth was observed without the following single amino acids: Ile, Leu, Thr and Tyr, although it occurred at later time points. Therefore, it was concluded that removal of these amino acids delayed growth, while lack of Arg, Cys, Trp or Val seemed to prevent growth up to time 48 hours and, therefore, these amino acids were considered 'essential' for growth. Simultaneous removal of more than two amino acids from the medium seemed to prevent growth in RP62A and all other strains tested.

## **1.6 Nitrogen assimilation and amino acid catabolism in staphylococci**

The mechanisms for N assimilation and the key aspects of amino acid synthesis and catabolism are important for understanding the functioning of a system and the metabolic capability of an organism.

### **1.6.1 Nitrogen assimilation**

The most common mechanisms of bacterial N assimilation involve the action of the enzyme Glt-dh (glutamate dehydrogenase) or the GS/GOGAT (Gln synthetase/Glt synthase) cycle. This has been

particularly well described for *E. coli*, which is known to primarily assimilate  $\text{NH}_4^+$  via the action of the Glt-dh (i), until the  $\text{NH}_4^+$  concentration falls below 1mmol, when its affinity for this substrate decreases and the GS/GOGAT system takes over (Amon *et al.* 2010) (ii). In summary, the mechanism works as follows (Somerville 2016): i) N is assimilated from  $\text{NH}_4^+$  by the Glt-dh catalysing a reductive amination of 2-KG that generates Glt; or ii) Glt (Tempest *et al.* 1970) is produced by the Glt synthase catalysing the transamidation of Gln and 2-KG that generates two molecules of Glt, while Gln can in turn be synthesised from Glt, ATP and  $\text{NH}_3$  by the Gln synthetase, or, alternatively, produced as glutamyl-tRNA by the action of a tRNA-dependent amidotransferase over glutamyl charged tRNA (Ito *et al.* 2010). Glt and Gln are central to amino acid biosynthesis, although it is Glt which serves as the primary amino donor in most biosynthetic reactions for amino acids (excluding Asn, Trp, and His) (Reitzer 2003).

## 1.6.2 The effect of glucose on amino acid catabolism

Despite amino acid catabolism being important for growth in different niches, little is known about it in staphylococci. Amino acid catabolism in staphylococci can be affected by the presence of Glc in the media (Townsend *et al.* 1996; Li *et al.* 2010; Nuxoll *et al.* 2012; Halsey *et al.* 2017). In 2017, Halsey *et al.* studied *S. aureus* growth in HHW, a rich but chemically-defined medium, by performing several growth analyses on mutants for catabolic pathways of interest selected from a sequenced transposon library and assessing the fate of  $^{13}\text{C}$ -labeled amino acids included in the media via NMR metabolomics (Halsey *et al.* 2017). They observed that when Glc was added at a concentration of 0.25% w/v (13.9 mmol), Glc, Glt, Asp and the glycogenic amino acids Ala, Ser, Gly and Thr were rapidly consumed and used to support growth, while catabolism of Arg, Pro and Lys was absent. On the other hand, cells grown in Glc-free medium utilised amino acids as follows: Glt acted as the main amino donor for anabolic processes as well as the main source of C by fuelling the TCA cycle via 2-KG and allowing subsequent gluconeogenesis via PEP synthesis from oxalacetate. Arg, Pro and His were directly transformed into Glt, without need for the involvement of Glt synthase (GOGAT) and the TCA cycle (Halsey *et al.* 2017; Halsey *et al.* 2017). While Pro acted as the main Glt source, His was only consumed upon depletion of Pro and Arg. Glycogenic amino acids were catabolised to Pyr, which was then consumed in the Pta/AckA pathway generating Ac and ATP via substrate-level phosphorylation during the exponential growth phase. An increased consumption of Ac was observed during the post-exponential growth phase, suggesting its transformation into AcCoA to enter the TCA cycle. Finally, Asp and Asn seemed to be mostly used to obtain oxaloacetate in order to replenish the TCA cycle, while other amino acids (Cys, Ile, Leu, Lys, Met, Phe, Tyr and Val) were consumed at a gradual pace, indicating their utilisation in the direct synthesis of protein.

In 1999, Tynecka *et al.* described the effect of Glc on Glt catabolism during the study of mechanisms for energy conservation in *S. aureus*: their results showed that starved cells supplemented with Glt

oxidised this to generate 2-KG and NADH, which respectively entered the TCA cycle and the respiratory chain via the NADH-menaquinone oxidoreductase leading to ATP synthesis by oxidative phosphorylation (Tynecka *et al.* 1999). However, if the medium was supplemented with 1% Glc, repression of the NAD-dependent Glt-dh and the 2-KG-dh stopped the organism from catabolising Glt. These results suggested that phosphorylation is tightly coupled with Glt oxidation when this is the main C source in the media but its catabolism is no longer a main metabolic strategy followed when Glc is present.

Glc was also shown to have an effect on the Ser catabolism of *S. epidermidis* (Sivakanesan *et al.* 1980). In 1980, Sivakanesan and Dawes showed how Pyr production via deamination of Ser was substantially decreased in an undefined rich medium when supplemented with a Glc concentration of 0.5% (w/v), both in aerobic and anaerobic conditions. This again suggests that, in general, amino acid catabolism in the presence of preferred C sources is strongly repressed in staphylococci.

### **1.6.3 Differential amino acid utilisation on planktonic cultures and biofilms**

Cells growing as part of biofilms are generally exposed to different biochemical conditions than planktonic cultures, such as lower concentration of nutrients and O<sub>2</sub> and a gradient of these concentrations across biofilm layers. It is, therefore, not surprising for cells to respond to these changes and to their need to produce specific biofilm biomass components by re-wiring their metabolism, which is likely to lead to specific amino acid extraction patterns. Such changes have been observed in staphylococci *in vitro*: particularly, a study conducted by Zhu *et al.*, in 2007 showed that *S. aureus* cultures growing as biofilms on TSB plus 0.25-0.5% w/v Glc selectively utilised the amino acids Arg, Gln, Gly, Pro, Ser and Thr while Trp and Val were left unconsumed (Zhu *et al.* 2007). In contrast, planktonic cultures, which also extracted Gln, Gly, Ser and Thr, consumed Ala and Glt, and took up only Pro and Val when other C sources became limiting. This study showed that although Arg utilisation seemed to favour biofilm formation in staphylococci by either i) helping to counteract excessive cell acidification via excretion of NH<sub>4</sub><sup>+</sup> via Arg deamination or ii) helping to avoid the cellular host immune response (Zhu *et al.* 2007), it was not essential for the process, since mutants showing reduced production of biofilm exopolysaccharides upon inactivation of the Arg deiminase operon *in vitro* were still able to generate biofilms as dense as the wild-type strains (Zhu *et al.* 2007).

In summary, the way in which Glc and other compounds are metabolised seems to vary substantially depending on the growth conditions (e.g. aerobiosis or anaerobiosis), the substrates present, the concentrations at which they are available and the growth phenotype and growth stage of the organism.

## 1.7 Biofilm production in staphylococci

The ability of NAS to colonize an implant surface by forming biofilms is the primary virulence factor that allows them to cause foreign-body related infections such as PJI (Arciola *et al.* 2012), (Tande *et al.* 2014). Biofilms are complex communities of bacteria embedded in an extracellular matrix. Bacterial cells encased in the biofilm matrix are difficult to eliminate, since they present reduced sensitivity to antibacterial and disinfectant agents and are partly sheltered from the host immune response (Arciola *et al.* 2012; Becker *et al.* 2014; Tande *et al.* 2014) Biofilms can be monomicrobial or polymicrobial (formed by one or more bacterial species) and are composed of bacterial subpopulations with different genotypic and phenotypic features (Arciola *et al.* 2012). Most biofilms involved in PJIs have a monomicrobial nature (Tande *et al.* 2014; Benito *et al.* 2016; Sebastian *et al.* 2018). Due to their structural and biological complexity, biofilms can be considered somehow similar to multicellular organisms (Arciola *et al.* 2012). Bacterial biofilms are mainly composed of water, which has been shown to account for up to 97% of the biofilm matrix when several bacterial biofilm types were analysed, followed by bacterial cells, exopolysaccharides and other excreted products (Sutherland 2001). A part from water, the biofilm matrix mainly contains extracellular polymeric substances (EPS), which are conglomerates of exopolysaccharides, proteins, lipids, teichoic acids and extracellular DNA (eDNA) (Arciola *et al.* 2012; Becker *et al.* 2014). The exact composition of a biofilm can vary depending on the environment where it is formed, the bacterial species involved and even between and within strains (Arciola *et al.* 2012; Tande *et al.* 2014; Zapotoczna *et al.* 2016). The ability of staphylococcal strains to form biofilms differs widely depending on environmental conditions and media composition (Arciola *et al.* 2012), and often, expression of biofilm formation *in vitro* has shown to be inconsistent. There is evidence that staphylococci are capable of adapting their metabolism to produce biofilms of different nature (proteinaceous, exopolysaccharidic, fibrinous or amyloid) according to external stimuli (Zapotoczna *et al.* 2016). It is also not clear whether the ability of an organism to form a luxuriant or a non-luxuriant biofilm could either be an advantage or disadvantage when colonizing a host (Arciola *et al.* 2012; Tande *et al.* 2014). When studying biofilm formation, staphylococci are commonly divided into *ica*-positive or *ica*-negative strains, depending on whether the *icaADBC* operon (Heilmann *et al.* 1996; Gerke *et al.* 1998; Lee *et al.* 2016; Somerville 2016) is either present or absent in their genome. As would be expected the biofilms vary, with *ica*-positive strains forming more structured and robust biofilms than the *ica*-negative strains (Le *et al.* 2018). The majority of the biofilm forming strains studied to date are *ica*-positive, as is the case with the strain selected to generate a metabolic model relevant to PJI in this project (*S. epidermidis* RP62A).

### 1.7.1 Molecular mechanisms of biofilm formation

Biofilm formation comprises several sequential events and is typically described as a four-step process (Arciola *et al.* 2012; Becker *et al.* 2014; Tande *et al.* 2014; Zapotoczna *et al.* 2016): (i) initial bacterial attachment to a surface, (ii) cell accumulation, (iii) biofilm maturation and (iv) cell detachment. There are diverse and redundant molecular mechanisms involved in biofilm formation, which can be summarized as follows: (Figure 1-4) (Arciola *et al.* 2012; Zapotoczna *et al.* 2016):

i) initial attachment:

i.i) adhesion through non-specific interactions (hydrophobic, electrostatic and Lifshitz-Van der Waals forces).

i.ii) adhesion through specific interactions. Mediated by autolysin/adhesin proteins (mainly AtlE in *S. epidermidis* and AtlA in *S. aureus*).

ii) cell accumulation:

ii.i) *ica*-positive strains: production of exopolysaccharides, including the main matrix component PIA (polysaccharide intercellular adhesine), for which synthesis is mediated by the *icaADBC* locus. Production of Microbial Surface Components Recognizing Adhesive Matrix Molecules (MSCRAMMs) such as the SdrG, SdrF, SesC or Embp adhesion proteins and production of phenol-soluble modulins, with a role in accumulation of amyloid aggregates which promote cell aggregation.

ii.ii) *ica*-negative strains: expression of adhesive proteins anchored to the cell wall such as Bap, Aap and Embp.

iii) biofilm maturation:

iii.i) increased production of components from the biofilm matrix (EPS). Once the cells reach stationary growth phase, their quorum sensing regulatory system (regulated by the *agr* and *lux* locus) activates, leading to an increased production of Autoinducing Peptide (AIP) and synthesis of the small RNA, RNAIII, a quorum sensing inhibitor (Giacometti *et al.* 2003).

iv) cell detachment:

iv.i) the increase on RNAIII leads to the expression of extracellular proteases and phenol-soluble modulins that cause cells to detach.

Overall, the regulation of bacterial biofilm formation is dependent on the detection of environmental stress and transcriptional regulation. In NAS, the alternative sigma factor SigB and the sar (staphylococcal accessory regulator) regulatory element SarA (Fluckiger *et al.* 1998) have been shown to be important, alongside the bacterial quorum sensing regulatory system.

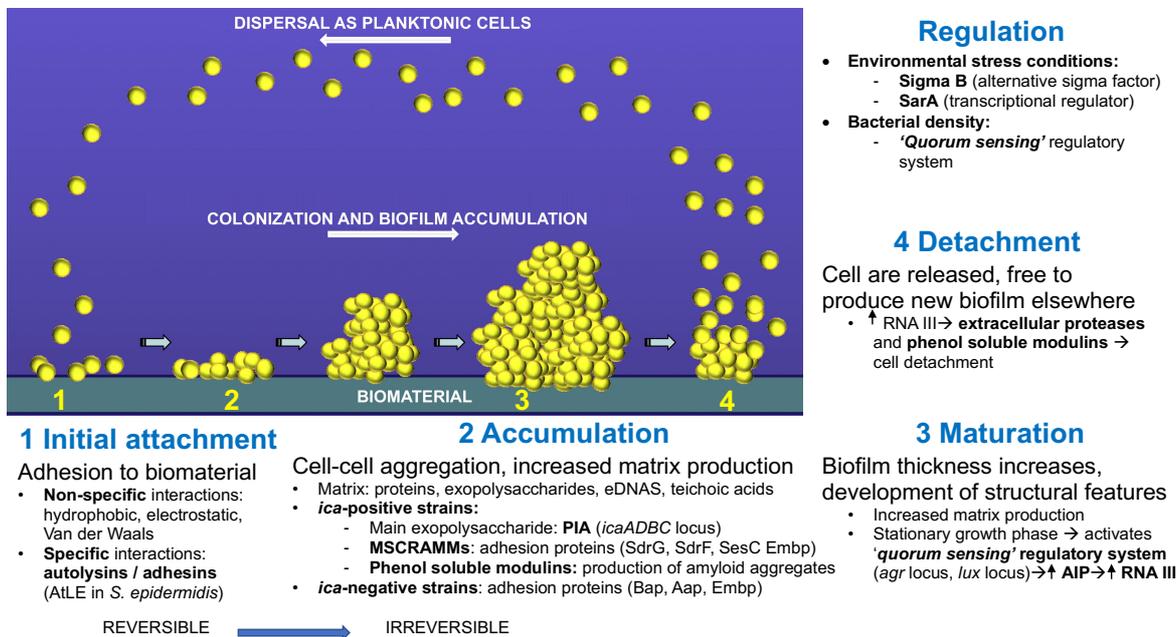


Figure 1-4 Schematic representation of the steps involved in the process of biofilm formation

Adapted from Arciola *et al.* 2012 and reproduced here with permission.

## 1.7.2 Environmental conditions encountered by bacteria growing in joints

### 1.7.2.1 Composition of the synovial fluid

The intra-articular space of the joints is filled with synovial fluid. This fluid is a mixture of a transudate of plasma and compounds actively secreted by synoviocytes (cells of the joint capsule) and its main role is to provide nutrients and lubrication for the cartilages. In normal physiological conditions, its composition is similar to that of plasma, with Glc levels close to those found in serum (550  $\mu\text{mol/l}$  approx.) and 25% of the total protein present in blood (Stein *et al.* 1954; Frame 1958; H Kenneth Walker 1990; Canepa *et al.* 2002; Gale 2007). Coagulation proteins are absent and small proteins (e.g. albumin) are present at higher concentrations than larger proteins such as globulins. Synovial fluid contains the same individual free amino acids as the human blood plasma, including non-essential amino acids such as aminobutyric acid, citrulline, ornithine and taurine. These range in concentrations from 2  $\mu\text{mol/l}$  (Asp) (Canepa *et al.* 2002) to 400  $\mu\text{mol/l}$  (Gln) (Canepa *et al.* 2002). Significant amounts of hyaluronic acid (16.65 mmol) (Gale 2007), a polymer of glucuronic acid and NAcGlc, are present and are responsible for the high viscosity of the fluid. Other components that can be found in this fluid are  $\text{NH}_4^+$ , chondroitin sulphate (a polymer of NAcGal and glucuronate),

urea and uric acid. While the concentration of  $\text{NH}_4^+$  is approximately 50  $\mu\text{mol/l}$  (Pathology Armony group 2011), the concentration of electron acceptors is approximately 70  $\mu\text{mol/l}$  for  $\text{NO}_3^-/\text{NO}_2^-$  (Tsikas *et al.* 1998; Ghasemi *et al.* 2010) and 24 mmol for  $\text{O}_2$  (Shapiro 1995; Chu *et al.* 2003; Malatesha *et al.* 2007). Since  $\text{NO}_2^-$  has a short half-life in plasma *in vitro*, being readily oxidised to  $\text{NO}_3^-$  (Gilchrist *et al.* 2010), the  $\text{NO}_3^-/\text{NO}_2^-$  blood concentration was solely attributed to  $\text{NO}_3^-$  in the corresponding *in silico* analyses performed during this project.

During inflammation and/or infection the following biochemical changes have been described (H Kenneth Walker 1990): decreased hyaluronic acid concentration and fluid viscosity due to its fragmentation by lysozymes released by polymorphonuclear cells (Fu *et al.* 2019), reduction on the Glc concentration due to bacterial activity, lower pH and, in some cases, an increase in lactic acid concentration. Thus, it is safe to assume that hyaluronic acid is likely to be found partially hydrolysed during inflammation, with the subsequent products (glucuronic acid and NAcGlc) being available in their free form.

### 1.7.3 Composition of biofilms growing in joints

To date, there is no biochemically defined biomass composition available for staphylococcal biofilms retrieved from prosthetic joints. As already mentioned, biofilm formation is a complex process and staphylococcal strains exhibit different strategies depending on their gene content and the environmental conditions encountered. RP62A is considered a reference biofilm-positive strain within *S. epidermidis* (Sadovskaya *et al.* 2004), the main species involved in nosocomial bacteraemia and indwelling medical device associated infections (Xue *et al.* 2015). *Ica*-positive strains, such as RP62A, are known to mainly (but not exclusively) generate biofilms of polysaccharidic nature. The main matrix component of these biofilms is a  $\beta$  1-6 linked NAcGlc polymer or PNAG (poly-NAcGlc), which has been shown to account for as much as 70% of the biofilm dry weight (Baldassarri *et al.* 1996). PNAG was first described in *S. epidermidis* (RP62A and other strains) and named Polysaccharide Intercellular Adhesin (PIA) (Mack *et al.* 1992; Mack *et al.* 1996). PIA is the major exopolysaccharide involved in intercellular adhesion in staphylococci (Somerville 2016) and plays an important role in the accumulation and maturation phases of biofilm formation in strains exhibiting an exopolysaccharidic biofilm-forming strategy (Büttner *et al.* 2015). It has been proposed as a marker for infection and even as a candidate for vaccine development (Somerville 2016). Several environmental factors have been reported to induce PIA synthesis, although with significant variation between strains (Otto 2008; Somerville 2016). These include: increased temperature; lack of  $\text{O}_2$ ; high osmolarity; presence of Glc and EtHO; and decreased iron concentration. Other important components of RP62A biofilms are extracellular teichoic acids, proteins and nucleic acids (Sadovskaya *et al.* 2004), however proportions in the biofilm biomass vary with the growth conditions. Both PIA and extracellular teichoic acids contain positive and negative charges in their structures (Somerville 2016). The ability of the cells to regulate the relative production of these

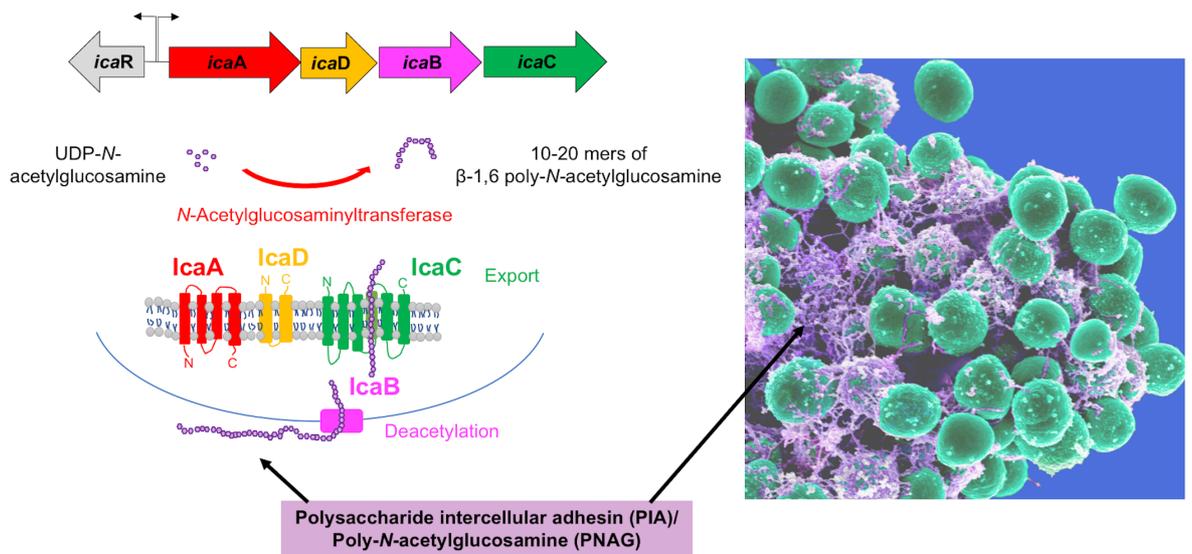
compounds allows them to generate biofilms with different physicochemical properties and, thus to colonise a wider range of environments (Sadovskaya *et al.* 2004).

#### 1.7.4 Structure and synthesis of PIA in *S. epidermidis*

The PIA polymer in *S. epidermidis* presents 15 to 20% of its sugar residues de-N-acetylated, thus positively charged, and distributed more or less evenly along the polymer (Mack *et al.* 1996). Approximately 6% of these residues are modified with succinate moieties, which introduce negative charges thus making the polymer zwitterionic (Mack *et al.* 1996). This characteristic is directly related to the adhesive properties of PIA, and its capacity to attach to the bacterial cell surface (Rohde *et al.* 2010), although the specific mechanisms by which this molecule is retained in the cell wall are currently unknown (Mack *et al.* 1996). The length of this polymer has been estimated in approximately 130 NAcGlc residues (Mack *et al.* 1996).

While the amino groups needed for PIA synthesis are obtained from amino acids, the production of sugar moieties is directly linked to glycolysis/gluconeogenesis: the amino sugar UDP-NAcGlc is the specific biosynthetic precursor and sugar donor for PIA production (Büttner *et al.* 2015; Lee *et al.* 2016; Somerville 2016). Synthesis of amino sugars involves directing Glc metabolism away from glycolysis and towards production of NAcGlc. In this process, an amino group is transferred to the intermediate F6P, generating GlcN-6P, which is subsequently isomerised to GlcN-1P. The last two steps in the *de novo* biosynthesis of UDP-NAcGlc involve the acetylation and activation of GlcN-1P and are catalysed by GlmU, a fused enzyme with two enzymatic activities (GlcN-1P acetyltransferase and UTP-*N*-acetyl- $\alpha$ -D-glucosamine-1-P uridylyltransferase) that has been identified as a potential therapeutic target for biofilm-associated infections (Burton *et al.* 2006; Somerville 2016).

In *S. epidermidis*, the *icaADBC* operon (Heilmann *et al.* 1996; Gerke *et al.* 1998; Lee *et al.* 2016; Somerville 2016) encodes all enzymes catalysing synthesis of PIA from UDP-NAcGlc, and has close orthologues in other species, including *E. coli*. The *ica* genes function as follows (Figure 1-5): *icaA* encodes for a NAcGlc transferase that generates an initial NAcGlc polymer of approximately 20 residues utilising UDP-NAcGlc as a substrate (Gerke *et al.* 1998; Somerville 2016; Le *et al.* 2018). This gene functions together with *icaD*, which assists the former in an undefined way. IcaC is a transmembrane protein involved in further elongation of the polymer and its transport toward the periplasm (Gerke *et al.* 1998; Somerville 2016; Le *et al.* 2018), and IcaB is a secreted deacetylase responsible for the de-N-acetylation step of some of the NAcGlc units, which introduces positive charges in the polymer by leaving the amino groups unmasked (Heilmann *et al.* 1996; Gerke *et al.* 1998; Somerville 2016; Le *et al.* 2018). The mechanism leading to O-succylation of the sugar residues is currently unknown.



**Figure 1-5 Schematic representation of the biosynthetic process of the biofilm matrix polymer PIA in staphylococci**

Figure from Le *et al.* 2018 and reproduced here with permission.

PIA, also known as PNAG, is a partially de-acetylated NAcGlc polymer involved in the accumulation and maturation phases of biofilm formation. It is synthesized by IcaA, a membrane-located NAcGlc transferase which works together with the accessory protein IcaD. IcaC is a transmembrane protein that exports the growing PIA chain, which is then partially de-acetylated by IcaB (a cell surface de-acetylase), introducing positive charges which are essential for the attachment of the polymer to the cell surface. The *ica* gene locus contains the *icaADBC* operon and the *icaR* gene, which respectively encode for the Ica proteins described above and the regulatory protein IcaR.

The regulation of PIA synthesis is fairly complex. There are multiple factors involved in the transcription of the *ica* operon, which can be regulated either directly at the *icaA* promoter or through expression of IcaR, which binds to the promoter region of the operon acting as a transcriptional repressor (Conlon *et al.* 2002). Some of the factors involved in this regulatory process are (Otto 2009; Cue *et al.* 2012): i) the action of global regulatory proteins, such as the sar family proteins, the global regulatory alternative sigma factor (SigB), the SrrAB two component system or the quorum sensing system; and, ii) the action of DNA insertion elements. While SarA and the alternative sigma factor SigB up-regulate PIA synthesis, the quorum sensing system luxS seems to down-regulate it (Otto 2008). The quorum sensing signalling molecule autoinducer-2 regulates biofilm formation in RP62A by regulating the *icaR* gene, although there are conflicting reports on whether its action up-regulates or down-regulates the *ica* operon (Xue *et al.* 2015). The widespread DNA insertion element IS256 can integrate into the *ica* genes abolishing PIA synthesis. This element can also integrate into the *sigB* operon and *sarA* locus, causing phenotypic switching and potentially affecting other cellular traits, allowing cells to adapt to the changing environment (O'Gara 2007). Its insertion seems to be irreversible (Otto 2008).

### 1.7.5 Metabolic features of staphylococcal biofilms

Despite their importance, there is very little consensus about the metabolic requirements for biofilm formation in staphylococci (Zhu *et al.* 2007): *S. aureus* biofilms have been reported to present a dynamic metabolic flux of C and amino acids (Zhu *et al.* 2007) while *S. epidermidis* biofilms have been described as maintaining a relatively low metabolic activity (Yao *et al.* 2005; Sadykov *et al.* 2010). However, studies of this area of *S. epidermidis* metabolism are scarce. It is believed that cells within biofilms grow in an anaerobic or microaerobic environment, where they convert Pyr to Lac, Ac and Form (Resch *et al.* 2005; Zhu *et al.* 2007), which leads to pH reduction and red-ox imbalances. These issues seem to be tackled by diverting Pyr to the butanediol pathway, where two pH-neutral compounds (acetoin and butanediol) are produced and excreted to the environment (Zhu *et al.* 2007), thus preventing excessive acidification (Section 1.4.4). Staphylococcal biofilms have also been reported to present specific amino acid consumption patterns (Section 1.6.3) and to redirect usage of C and energy from biosynthetic pathways towards growth support (Zhu *et al.* 2007). Amino acid catabolism and other aspects of biofilm metabolism are yet to be fully understood. This is partly due to the existence of several biofilm forming strategies across and within bacterial species, and partly because of the complexity of biofilm structures and their life cycles.

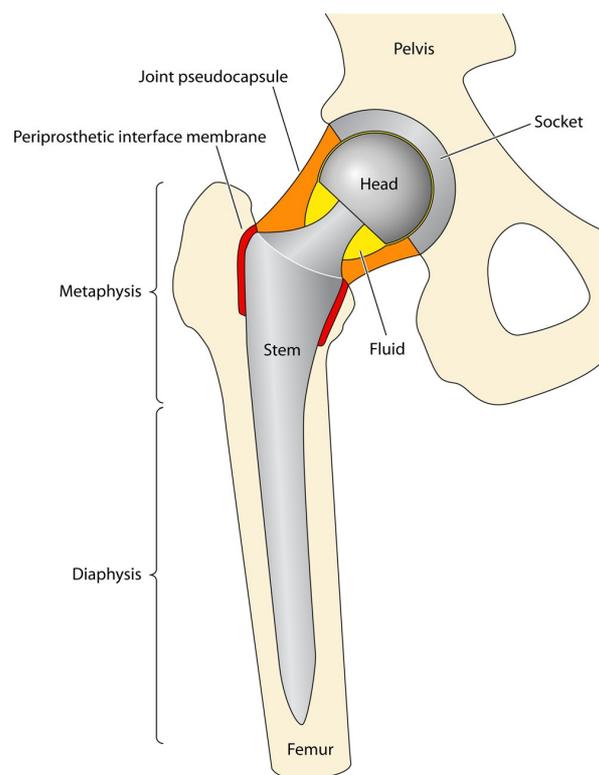
A general feature of exopolysaccharidic staphylococcal biofilms is that external conditions or stresses that repress central metabolism such as high Glc concentrations or restrictive levels of nutrients, iron or O<sub>2</sub> have been shown to result in dramatic increases of PIA production *in vitro*, hence, suggesting that changes on the intercellular levels of biosynthetic intermediates, ATP or the redox status of the cell may act as intracellular metabolic signals that influence the transcriptional processes regulating biosynthesis of PIA (Vuong *et al.* 2005; Sadykov *et al.* 2008; Sadykov *et al.* 2011; Somerville 2016). Thus the activity of the TCA cycle has been proposed as a mechanism for translation of external environmental stimuli into metabolic effectors with a final regulatory role on biofilm formation (Sadykov *et al.* 2008).

Further studies are needed in order to understand biofilm metabolism as a whole. Particularly, some authors are applying metabolic modelling to the investigation of the spatio-temporal distribution of substrates and their utilisation within these communities (Phalak *et al.* 2016; Carlson *et al.* 2018; Zhang *et al.* 2018; Schepens *et al.* 2019), which is an important step forward for those trying to unravel the metabolic features that make biofilms the most ubiquitous and successful colonising bacterial phenotype.

## 1.8 Prosthetic Joint Infection

Arthroplasty or joint replacement is a surgical procedure during which a damaged joint is realigned, remodelled or replaced by a prosthetic implant. It allows millions of patients to relieve pain and

recover or improve joint functionality each year (Tande *et al.* 2014). PJI occurs when bacteria manage to adhere to the surface of a prosthetic implant and colonize it by growing a biofilm. PJIs are usually acquired at the time of surgery through contamination of the prosthesis or periprosthetic tissue, or, more rarely, through a hematogenous infection (Tande *et al.* 2014). Observations in animal models suggest that, right after intraoperative contamination of the joint space the infection is confined to this area. Once established, it spreads to the adjacent bone metaphysis and, if allowed to progress, it could also affect the bone diaphysis (Belmatoug *et al.* 1996; Tande *et al.* 2014) (Figure 1-6). It is unclear if this is also the case in hematogenous infections (Belmatoug *et al.* 1996; Tande *et al.* 2014). Depending on the virulence of the causative organism, the symptoms might show up just a few months after surgery (< 3 months), between 3 and 12 months, or even 24 months later. PJI by NAS can occur in any of these three modalities (Tande *et al.* 2014).



**Figure 1-6 Schematic showing a total hip arthroplasty in place with relevant structures highlighted**

Figure from Tande *et al.* 2014 and reproduced here with permission.

### 1.8.1 The scope of the problem

PJI occurs with a relatively low frequency in primary hip and knee arthroplasties (1.5-2.5%) (Montanaro *et al.* 2011; Arciola *et al.* 2012) but it can have devastating consequences for the patients. Complications of this process can cause severe bone and soft tissue damage (Montanaro *et al.* 2007), (Arciola *et al.* 2011; Montanaro *et al.* 2011; Arciola *et al.* 2012). Often, PJI is diagnosed once the infection has already become chronic. At this point, substitution of the implant under revision surgery is the only option for successful treatment (Arciola *et al.* 2012; Lange *et al.* 2012) and the risk of relapse increases (3.2-5.6%) (Montanaro *et al.* 2011; Arciola *et al.* 2012). The economic impact of

PJI is also significant. Arthroplasty is a frequently performed procedure, and, due to the cumulative effect of an aging population and an increasing life expectancy presents a rising incidence, and so do PJIs. In 2015, there were over 185,000 primary joint replacements performed in the UK and 15,027 revision hip or knee surgeries. From those, 2,811 procedures were for management of PJI (National Joint Registry for England 2016). In Europe, the mean total costs per patient rose from €14,135 (between 1997 and 2001) to €23,113 (between 2002 and 2006) (Oduwole *et al.* 2010) and in the UK, the mean cost of a total knee arthroplasty for infection had already reached £30,011 per patient by 2012 (Kallala *et al.* 2015). These figures illustrate how PJI is becoming an increasing burden for the healthcare system and the population.

### 1.8.2 Etiological causes

The majority of PJIs are monomicrobial infections. Polymicrobial infections have been reported in up to 35% of early-onset cases and in less than 20% of cases at any time point after joint replacement (Tande *et al.* 2014; Benito *et al.* 2016; Sebastian *et al.* 2018). Staphylococci are the leading etiological agents (Arciola *et al.* 2012; Tande *et al.* 2014; Bémer *et al.* 2016; Benito *et al.* 2016). Among them, *S. aureus* and *S. epidermidis* respectively represent the first and the second bacterial species most often isolated in cultures from these patients (Arciola *et al.* 2012; Tande *et al.* 2014; Bémer *et al.* 2016; Benito *et al.* 2016). In 2014, Tande and Patel conducted a review of the published data regarding the relative frequency of microorganisms isolated from patients with PJI (Tande *et al.* 2014). They performed a collective analysis of the microbiological results of 14 large studies that included over 2,400 patients from several countries and time points (Table 1-1). Their observations concluded that staphylococci accounted for between the 50 and 60% of the cases, with *S. aureus* and NAS equally contributing to this percentage. Two later multicenter studies, performed by Benito *et al.* (Benito *et al.* 2016) in Spain and Bemmer *et al.* (Bémer *et al.* 2016) in France, presented similar results, with staphylococci being the leading cause of PJI (in 65.2% and 66.3% of the cases respectively). In the datasets from the 2014 review study (Tande *et al.* 2014) and the French study (Bémer *et al.* 2016), *Streptococcus* and *Enterococcus* species accounted approximately for the 10% of the cases and aerobic Gram-negative bacilli for less than 10%, with both studies reporting approximately a 15% of polymicrobial infections, while the Spanish study (Benito *et al.* 2016) reported an increasing trend in infections caused by Gram-negative bacilli and the rise of multidrug-resistant infections (mainly due to this group of organisms).

**Table 1-1 Common etiological causes of PJI**

Infection	% of patients with prosthetic joint infection					
	Hip and knee					
	All time periods <sup>a</sup>	Early infection <sup>b</sup>	Hip <sup>c</sup>	Knee <sup>c</sup>	Shoulder <sup>d</sup>	Elbow <sup>e</sup>
<i>Staphylococcus aureus</i>	27	38	13	23	18	42
Coagulase-negative <i>Staphylococcus</i>	27	22	30	23	41	41
<i>Streptococcus</i> species	8	4	6	6	4	4
<i>Enterococcus</i> species	3	10	2	2	3	0
Aerobic Gram-negative bacilli	9	24	7	5	10	7
Anaerobic bacteria	4	3	9	5		
<i>Propionibacterium acnes</i>					24	1
Other anaerobes					3	0
Culture negative	14	10	7	11	15	5
Polymicrobial	15	31	14	12	16	3
Other	3					

<sup>a</sup> Data aggregated from 2,435 joints (24, 26, 57, 79, 93–102).

<sup>b</sup> Data aggregated from 637 joints (67, 97, 98, 103–107).

<sup>c</sup> Data from 1,979 hip and 1,427 knee PJIs from the Mayo Clinic Prosthetic Joint Infection Database (E. F. Berbari, personal communication).

<sup>d</sup> Data aggregated from 199 shoulders (56, 110–116).

<sup>e</sup> Data aggregated from 110 elbows (13, 117–120).

Data from Tande *et al.* 2014 and reproduced here with permission.

### 1.8.3 Diagnosis

Orthopaedic surgeons have to make a diagnostic decision based on the interpretation of the symptoms of the patients, their clinical history, the results of clinical investigations and the microbiological results obtained by the clinical laboratory (Atkins *et al.* 1998). *S. aureus* is usually far more virulent than NAS and its detection on a patient's samples is sufficient to give a positive diagnosis for infection. However, NAS are common skin commensals and comprise an important source of sample contamination and, therefore, their isolation leads to diagnostic uncertainty (Tande *et al.* 2014). Most clinical laboratories do not type NAS but simply report as "coagulase-negative staphylococci". The decision on clinical relevance is based on the repeated isolation of NAS with similar antibiotic resistance profiles and the clinical picture of the patient (Parvizi *et al.* 2011; Becker *et al.* 2014).

#### 1.8.3.1 Typing of NAS

Accurate identification of NAS to the species level is difficult (Ghebremedhin *et al.* 2008). During the last two decades, advances on next generation sequencing techniques have led to the development of new typing schemes for these organisms. Several typing methods use sequencing data from the variable regions of the multi-copy 16S rRNA gene (Takahashi *et al.* 1999; Becker *et al.* 2004), specific regions of housekeeping genes, such as *sodA*, *tuf*, or *gap* (Ghebremedhin *et al.* 2008), whole genome sequencing (Becker *et al.* 2014) or multi-locus sequencing typing (Miragaia *et al.* 2007). So far, no definitive typing scheme has been agreed.

Clinical laboratory speciation of NAS isolates is typically based in biochemical testing (e.g. gram tests, catalase or coagulase activity), antibiogram profiling or, the more recently introduced MALDI-TOF MS (Carbonnelle *et al.* 2007; Zhu *et al.* 2015), which is the current gold standard method for routine identification of staphylococcal species (Argemi *et al.* 2015). It identifies bacterial species

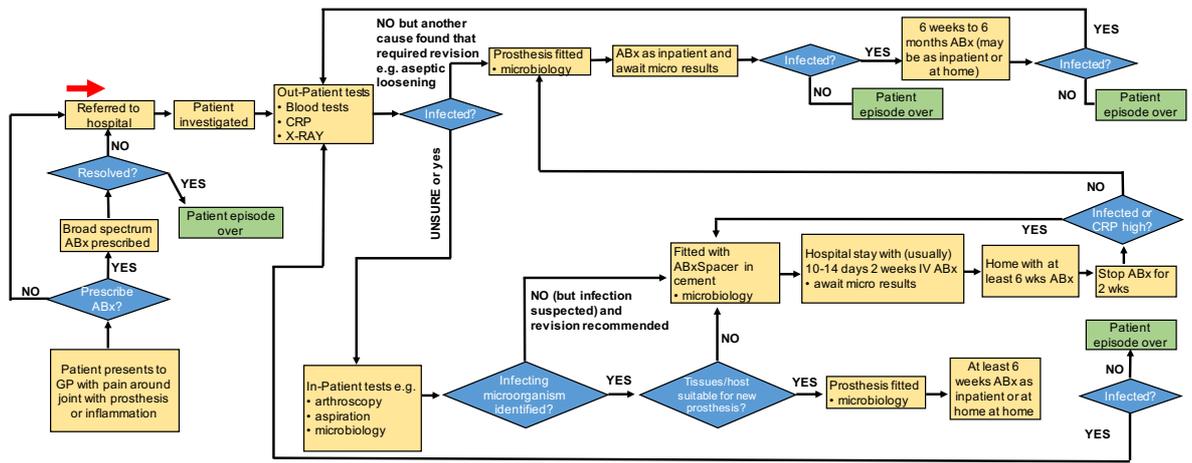
by comparing the peptide mass spectra of the isolates with stored profiles derived from the current definition of NAS species. However, there is some doubt regarding the taxonomical and biological relevance of the current phylogenetic classification of NAS (Section 1.3.2).

### ***1.8.3.2 Clinical diagnosis: the decision-making process***

The existence of three different diagnostic guidelines (English, French and American), (de Languue Française 2010; Osmon *et al.* 2012; Public Health England 2016) evidences the difficulty of diagnosis for clinicians. Despite being based upon data provided by the same microbiological study (Atkins *et al.* 1998), these guidelines differ in their recommendations regarding the collection and processing of intraoperative biopsies and on how to assign clinical significance to the isolates. In general, a patient is considered positive for infection if multiple ‘indistinguishable isolates’ (2 to 3) are identified on several fluid or tissue samples (3 to 5). Two isolates are defined as ‘indistinguishable’ if they share similar antibiograms. However, different strains may present similar antibiograms and genetically indistinguishable strains may present different ones, so this definition can be ambiguous. The current UK guidelines are published by Public Health England (Public Health England 2016)

### **1.8.4 Treatment**

PJI caused by NAS can present ambiguous symptoms such as moderate pain, swelling of the joint and reduced range of movement, which could also be compatible with a presumed non-infectious process known as aseptic loosening (Tande *et al.* 2014). The treatment for PJI and aseptic loosening differs significantly on its length and level of aggressiveness and has, as a consequence, different implications for the patient and different associated costs to the healthcare system (Figure 1-7). Usually, treatment of aseptic loosening includes a single stage revision surgery and an oral antibiotic regime (Tande *et al.* 2014; Norfolk & Norwich University Hospital 2015). In contrast, treatment for PJI involves surgical debridement or implant revision of the joint (in many cases, a two-stage surgery), which carries significant surgical morbidity and requires high doses of antibiotics between the two stages (usually for around six weeks) and after surgery (Osmon *et al.* 2012; Becker *et al.* 2014; Norfolk & Norwich University Hospital 2015). Often, the causative organism cannot be identified prior to surgery, leading to unnecessarily aggressive treatment of instances where symptoms are derived from causes other than an infection. Therefore, making an accurate diagnosis represents a major decision for the patient and the surgeon (Becker *et al.* 2014).



**Figure 1-7 Knee revision at Norfolk & Norwich University Hospital showing actions and decision points**

Actions are shown in boxes, decision points in diamonds, diagnostic test-related activities are shown in bullet points and the use of antibiotics appears as ABx. This figure was produced by Iain McNamara and colleagues (Norfolk & Norwich University Hospital) and is reproduced here with permission.

## 1.9 Metabolic reconstructions of staphylococci

Mathematical analysis of metabolic networks combines the use of bioinformatics tools, mathematical theories and *in silico* simulation as a way to explore and study cellular behaviour (Feist *et al.* 2009). It allows the integrative analysis of a wide range of experimental data (Shlomi *et al.* 2007; Oberhardt *et al.* 2009) and has been successfully used over the years to test hypotheses, simplify and increase the accuracy of the experimental design process (Feist *et al.* 2009; Oberhardt *et al.* 2009), define strategies for the genetic manipulation of organisms with commercial purposes (Fong *et al.* 2005; Durot *et al.* 2009; Kabimoldayev *et al.* 2018) and to define new drug targets (Becker *et al.* 2005; Jamshidi *et al.* 2007; Raghunathan *et al.* 2009). A more detailed introduction to the mathematical modelling of metabolism is provided in the following chapter.

At the time of writing, a curated genome-scale metabolic model of *S. epidermidis* has not been published, therefore, this project is expected to have a high research impact. However, several models have been published for *S. aureus*, most of them for the strain N315, and these have been used to define novel drug targets against staphylococci (Becker *et al.* 2005; Heinemann *et al.* 2005), new broad-spectrum drug targets (Lee *et al.* 2009) and even to develop new typing schemes (Bosi *et al.* 2016):

Two initial models, named iSB619 (Becker *et al.* 2005) and iMH551 (Heinemann *et al.* 2005), were published in 2005. Both studies used linear programming-based analysis to compute the effects of gene deletion in the presence of O<sub>2</sub> and rich media, and defined essential reactions for growth under these conditions (Becker *et al.* 2005; Heinemann *et al.* 2005). The first model (iSB619) was used to define an *in silico* minimal medium for bacterial growth, which compared reasonably well with an experimentally defined medium, but presented some discrepancies regarding amino acid requirements. This work identified a list of possible chemical inhibitors for 24 reactions defined as

essential for growth and were proposed as candidates to be considered in the search for new anti-staphylococcal compounds (Becker *et al.* 2005). The second model (iMH55), was used to perform an *in silico* gene deletion analysis that defined the sub-metabolism of glycans, lipids, cofactors and vitamins as fairly rigid and proposed for them to be considered as potential new drug targets.

The two latest models (by Lee *et al.* 2009 and Bosi *et al.* 2016) were initially developed for the strain MRSA N315 but were later modified to include several other strains (12 and 63 respectively). The model published by Lee *et al.* was used to define an *in silico* minimal medium for *S. aureus* that contained 6 components less than a previously published minimal medium (Price *et al.* 2004). Experimental results showed that both media supported similar growth rates *in vitro* (Price *et al.* 2004; Lee *et al.* 2009). These authors also proposed studying enzymatic genes as strong candidates for new drug targets (Lee *et al.* 2009), since they tend to present less variation than genes involved in pathogenicity or virulence. An enzyme deletion study performed with this model defined several single and paired metabolic enzymes as unconditionally essential for growth (in rich media) (Lee *et al.* 2009). The results obtained indicated that the functionality of the metabolic networks of these strains was strongly strain dependant. The work published by Bosi *et al.* in 2016 determined the pan-genome (total set of genes) and the core genome (genes shared by all the strains) of a total of 64 strains (Bosi *et al.* 2016). The 58% of the genes in the core genome were found to have metabolic functions. The virulome (total set of genes encoding virulence factors) was also defined, with the 39% of these genes shared by all strains. Finally, comparison of these models showed that the core reactome (reactions shared by all strains) represented, on average, the 70% of all reactions in each independent strain. Reactions which differed between strains were mainly involved in the catabolism of alternative C sources and amino acid biosynthesis, and therefore, possibly related to niche adaptation. The models were used to analyse the strain's ability to utilize up to 300 different nutrients in the presence and absence of O<sub>2</sub>. Finally, data regarding the strain's growth capabilities and distribution of virulence factors was combined and proposed as a way to define new typing schemes with potential to distinguish disease phenotypes (Bosi *et al.* 2016).

A detailed comparison of the general properties of these models and the model produced during this project is given in Chapter 7, Section 7.1.

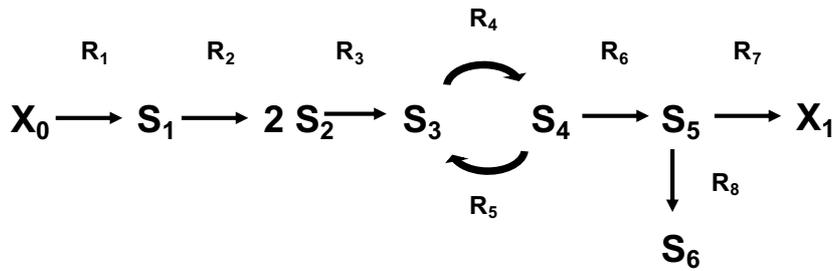
## 2 General methods

### 2.1 Introduction to mathematical modelling of metabolism

Mathematical modelling builds conceptual representations of real-world processes using mathematical language and techniques as a way to facilitate their study. Mathematical models can be used to explore the characteristics of a system: representing a metabolic network in a mathematical format makes it accessible to computers for extensive and rigorous analysis. Models can be built to study either small sections or the full metabolic network of an organism of interest (Poolman *et al.* 2013) and can be used to perform simulations to help answering complicated questions that would otherwise be too time-consuming or expensive to investigate with an experimental approach (Hofmeyr 1986). Some of its applications include *in silico* hypothesis testing, guiding experimental design (Ives *et al.* 1997; Feist *et al.* 2009; Oberhardt *et al.* 2009) interrogation of multi-species relationships (Phalak *et al.* 2016; Carlson *et al.* 2018; Zhang *et al.* 2018; Schepens *et al.* 2019), network property discovery (Edwards *et al.* 1999) or the identification of novel drug targets and virulence factors (Becker *et al.* 2005; Jamshidi *et al.* 2007; Raghunathan *et al.* 2009).

Metabolic networks can be defined as sets of reactions in which metabolites are either produced or consumed. The rate in which the concentration of a metabolite changes is equal to the difference between its production and consumption in the network and is, therefore, subjected to mass balance constraints. This is the fundamental basis of any metabolic model. The structure of a network can be represented mathematically in the form of a stoichiometry matrix storing information about the relationship between network reactions and their associated metabolites and will be described in more detail in the following section. The metabolites in a network can be considered either internal, if no net change in their concentration occurs within the system, or external, if their concentration is maintained at a constant level outside the system (Heinrich *et al.* 1996) (Figure 2-1). External metabolites are likely to be in permanent exchange across the system's boundary.

Figure 2-1 below illustrates a simple hypothetical network where the letter 'S<sub>n</sub>' is used to represent the concentration of internal (or variable) metabolites while 'X<sub>n</sub>' represents that of external (and 'clamped' or concentration-invariable) metabolites.



**Figure 2-1 Schematic representation of a simple set of reactions**

There are two main types of models representing metabolic networks: structural and kinetic (or dynamic) metabolic models. While structural models study a system during an specific moment in time and take into account the stoichiometry and thermodynamics of metabolic reactions, kinetic models study how systems behave over time and, therefore, also include parameters regarding reaction kinetics. Construction of dynamic models requires determining kinetic parameters associated with each reaction and the intracellular concentration of the metabolites involved. Obtaining these data requires a great effort and so far, still not possible for networks containing several hundreds or thousands of reactions, as is the case of genome-scale networks. Furthermore, their mathematical analysis requires a much higher computational power. For these reasons, most genome-scale models are structural models while kinetic models are generally restricted to smaller networks (Poolman *et al.* 2013). Structural metabolic modelling is the type of modelling used for the purpose of this project.

## 2.2 Structural metabolic modelling

Structural metabolic models and their analysis depends on the following assumptions (Singh *et al.* 2015):

- i) The metabolites in the system can be classified as concentration variable metabolites (internal) or metabolites which concentration remains unchanged independently of the actions taken by the system (external).
- ii) The rate in which the concentration of internal metabolites changes equals the sum of their rate of production and consumption.
- iii) The system can reach the steady state, where there is no net production or consumption of internal metabolites.
- iv) The reactions in the system can be classified as reversible or irreversible.

The rates of change in the concentration of metabolites involved in these reactions can be expressed as the following set of differential equations, where ' $S_x$ ' represents metabolite concentrations;  $t$  represents time and ' $v_x$ ' represents reaction fluxes:

$$\begin{aligned}
 dS_1/dt &= v_1 - v_2 \\
 dS_2/dt &= 2v_2 - v_3 \\
 dS_3/dt &= v_3 - v_4 + v_5 \\
 dS_4/dt &= v_4 - v_5 - v_6 \\
 dS_5/dt &= v_6 - v_7 - v_8 \\
 dS_6/dt &= v_8
 \end{aligned}$$

In structural modelling, the stoichiometry of the system can be represented as a stoichiometry matrix, conventionally denoted as  $N$ : in this matrix, reactions are represented as columns and metabolites as rows, with each entry corresponding to the stoichiometric coefficient for a given metabolite in each given reaction, where positive values denote metabolite production, negative values denote consumption and a value of 0 indicates no involvement of the metabolite in the reaction (Hofmeyr 1986; Heinrich *et al.* 1996; Poolman *et al.* 2007). This way, the structure of the metabolic network described in Figure 2-1 can be expressed as a stoichiometry matrix which compiles the stoichiometries of all reactions in the network ( $R_1$  to  $R_8$ ):

$$N = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The set of equations presented above can be represented in a matrix notation where the variation in the concentration of metabolites over time can be expressed as a result of multiplying the stoichiometry matrix ( $N$ ) by a vector of reaction fluxes ( $v$ ), as shown in the equation below (Equation 2-1):

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} v_{R1} \\ v_{R2} \\ v_{R3} \\ v_{R4} \\ v_{R5} \\ v_{R6} \\ v_{R7} \\ v_{R8} \end{bmatrix} = \begin{bmatrix} dS_1/dt \\ dS_2/dt \\ dS_3/dt \\ dS_4/dt \\ dS_5/dt \\ dS_6/dt \end{bmatrix}$$

### 2.2.1 The steady state assumption

A system of metabolic reactions can be considered to remain in the steady state, hence without net changes in the concentration of internal metabolites. Taking into account the assumptions ii) and iii)

defined above, the system of differential equations derived from the reaction network described in Figure 2-1 can now be re-defined as follows:

$$\begin{aligned} dS_1/dt = 0 &= v_1 - v_2 \\ dS_2/dt = 0 &= 2v_2 - v_3 \\ dS_3/dt = 0 &= v_3 - v_4 + v_5 \\ dS_4/dt = 0 &= v_4 - v_5 - v_6 \\ dS_5/dt = 0 &= v_6 - v_7 - v_8 \\ dS_6/dt = 0 &= v_8 \end{aligned}$$

At steady state, the net production of the variable metabolites is balanced with their net consumption and information about the relationships of reaction rates can be obtained by analyzing the system mathematically (Schuster *et al.* 2007).

If described in a matrix notation, the system of differential equations described above can now be expressed as the following equation (Equation 2-2):

$$N * v = 0$$

where  $N$  represents the stoichiometry matrix,  $v$  represents the vector of reaction rates or fluxes and  $N * v = 0$  denotes steady state conditions. This equation is under-determined, hence there is an infinite number of solutions that satisfy it.

### 2.2.2 Null space analysis

Since differential equations derived from a metabolic network are linear functions of reaction fluxes, solutions for the system of equations can be generated using null space analysis (Hofmeyr 1986; Poolman *et al.* 2007) by exploring solutions for Equation 2-2, thus the mathematical analysis of structural models is based on finding instances of  $v$  that satisfy  $N * v = 0$ . The  $v$  vector is not unique, since there would be infinite number of  $v$  vectors that satisfy this equation. Structural analysis identifies individual instances of  $v$  that satisfy this equation and the properties that any  $v$  vector must have for any steady state. Null space analysis identifies the invariant properties of all possible instances of  $v$  applying standard techniques of linear algebra to the identification of a null space or kernel matrix for  $N$ , conventionally denoted as  $K$ . Any set of steady state fluxes can be represented as a linear combination of the columns in the kernel. In  $K$ , reactions are represented as rows while the number of columns denotes the dimension of the solution space containing all possible steady states. Hence, the kernel matrix corresponding to the network represented in Figure 2-1 is:

$$K = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 0 & -1 \\ -1 & 0 \\ 1 & -1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}$$

The investigation of solutions for Equation 2-2 leads to the identification of the following invariant properties of the network:

- i) **Dead reactions**, or reactions which row vector in  $K$  contains solely zero elements. These reactions cannot carry flux in the steady state. For example, by calculating  $K$ , it is possible to readily identify reaction  $R_8$  in the metabolic network represented in Figure 2-1, with row vector  $[0\ 0]$ , as a dead reaction which could not carry flux in any steady state solution.
- ii) **Enzyme or reaction subsets** (Pfeiffer *et al.* 1999), or sets of reactions that must carry flux in a fixed ratio in any steady state solution. The non-zero elements in the row vectors of  $K$  for these reactions are scalar multiples of each other and, therefore, keep a constant ratio. Thus, reactions  $R_1$ ,  $R_2$ ,  $R_3$ ,  $R_6$  and  $R_7$  in Figure 2-1, which all share the same  $K$  row vector independently of the coefficients sign ( $[\pm 1\ \pm 1]$ ), form an enzyme or reaction subset, since all must carry flux in a fixed proportion in the steady state.

Any feasible set of reaction fluxes at steady state will be a linear combination of the matrix kernel, which forms the basis for the null space of the stoichiometry matrix. It is important to keep in mind that null space analysis does not provide an unique solution for the system, but defines the space containing all possible solutions. It also does not take into account thermodynamic considerations and, therefore, can be misleading when identifying the impact of enzyme deletions.

Although in metabolic modelling the term null space generally refers to the right null space of the stoichiometry matrix ( $N$ ) containing all steady-state flux solutions, the left null space reflects the moiety conservation relationships that are true at all states, not just the steady-state and corresponds to the right null space or kernel of the transpose of  $N$ . Moieties presenting the same coefficients in the left null space matrix are conserved moieties. Further study of the left null space is out of the scope of this project.

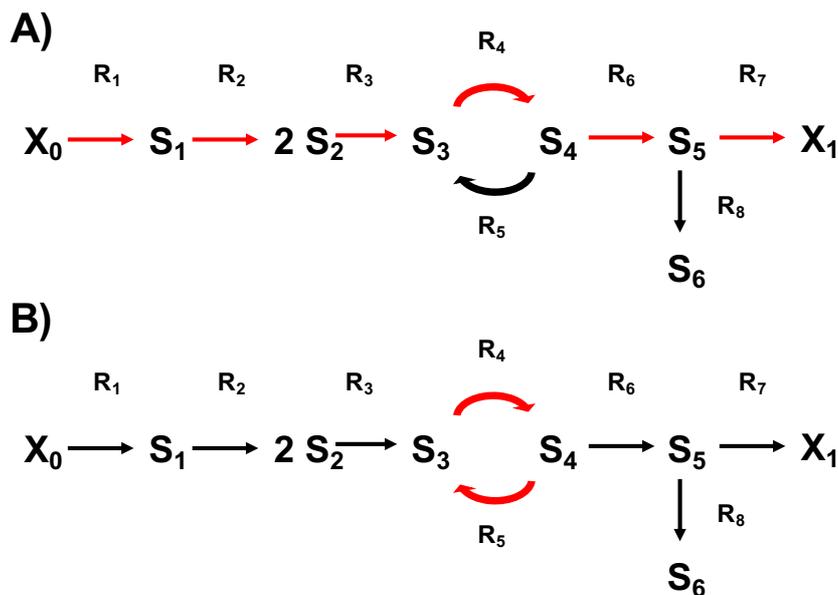
### 2.2.3 Elementary modes analysis

Elementary modes are minimal steady-state routes through the system which cannot be decomposed further, with all reactions involved working in the thermodynamically-favored manner. The concept of elementary modes was first proposed by Schuster *et al.* in 1999 and was achieved by exploiting

the fact that columns on the  $K$  matrix represent minimal independent routes through the network (although they do not take into account the reversibility of reactions) (Schuster *et al.* 1999). While determining the kernel of a stoichiometry matrix is a computationally simple task, calculation of elementary modes is subjected to combinatorial explosion and, therefore, it is conventionally applied to networks of small size.

Calculating the elementary modes of a network checks for the existence of feasible routes through the system. Since each elementary mode has an overall net stoichiometry, they can be used to identify the most productive pathways for production of compounds of interest. For example, this type of analysis has been used before in *E. coli* to identify the most efficient routes for production of by-products such as DAHP (Trinh *et al.* 2008) and ethanol (Van Dien *et al.* 2002); or to obtain a deeper understanding of the metabolic pathways occurring in mitochondria (Nicolae *et al.* 2015) etc.

This way, applying elementary modes analysis to the metabolic network described in Figure 2-1 two independent minimal routes through the system were defined (Figure 2-2):



**Figure 2-2 Elementary modes identified in the metabolic network described in Figure 2-1**

Reactions involved in the elementary mode are represented in red. A): elementary mode converting the external metabolite  $X_0$  into the external metabolite  $X_1$ . B) Elementary mode representing an internal metabolic cycle where  $S_3$  and  $S_4$  are interconverted without net production or consumption of either, thus this mode has no net stoichiometry.

## 2.2.4 Linear programming-based analysis

Another type of mathematical analysis that can be applied to structural metabolic models is linear programming-based (LP-based) analysis, which identifies instances of  $v$  that satisfy the steady-state

assumption ( $N \cdot \mathbf{v} = \mathbf{0}$ ) for a given objective function, while also satisfying additional constraints defined by the user and thus identifies specific flux patterns through the network (Fell *et al.* 1986; Edwards *et al.* 2002). During LP-based analysis, the steady-state constraint is maintained while a specific objective is defined, such as maximizing the production of biomass per unit of substrate consumed (Varma *et al.* 1994; Schuster *et al.* 2008) or minimizing the total net flux through the system, which has been utilized as a proxy for reducing the enzymatic cost of a metabolic process (Holzhutter 2006; Poolman *et al.* 2009) by assuming flux through the network to be equivalent to enzymatic activity. These correspond to hypothetical biological objectives followed by the organism from which the model is derived. Also, further constraints are applied to single or multiple reaction fluxes by fixing the value, or defining the range of values, these fluxes can take. This is useful to, for example, explore the behavior of the system under specific conditions (e.g. in the presence or absence of certain substrates or electron acceptors) or to study specific processes, such as production of ATP or production of certain by-products from certain substrates. An example of a generic LP formulation is described below (Equation 2-3):

$$\begin{array}{l} \text{Minimize or maximize: } \mathbf{V}_{target} \\ \\ \text{subject to } \left\{ \begin{array}{l} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ \\ \min \leq \mathbf{v}_R \leq \max \end{array} \right. \end{array}$$

where  $V_{target}$  represents a target reaction, being the objective of the LP formulation to either minimize or maximize flux through this reaction;  $\mathbf{v}$  is the vector of reaction fluxes and  $N$  is the stoichiometry matrix, with  $n$  rows (metabolites) and  $m$  columns (reactions), therefore, constraining the LP to obey the steady state assumption, while  $V_R$  defines the flux through a selected reaction, being constraint between a lower (*min*) and an upper (*max*) flux bound.

The objective value of a LP-based analysis solution provides valuable information. For example, if the objective of the analysis is to minimise the total net flux through the system, a higher objective value indicates that the network needs to invest more flux to satisfy the constraints imposed in the analysis. This value can be used to compare the efficiency of solutions obtained upon different constraints.

Each solution obtained from solving a LP problem is an optimal flux distribution for the given objective, but it is not unique. Multiple optimal and sub-optimal solutions might exist. Furthermore, the information provided by a single solution gives very limited insight into the system's behavior and the way reaction fluxes might respond in order to adapt to the changing conditions or to fulfill certain demands. The same LP problem can be repeatedly solved by varying a single constraint at a

time or by sequentially incrementing one or more constraints, thus ‘scanning’ the model under constraints that mimic different environmental or physiological conditions. This way, a set of solutions can be obtained and used to identify coordinated responses (e.g. reactions that respond to a sequential increase in the ATP demand). This allows for a better understanding of the range of metabolic responses that can be exhibited by the system and, therefore, by the organism it represents.

## **2.3 Software and metabolic modelling tools**

### **2.3.1 The Python programming language**

Python (M. Lutz 1999; Lutz 2001) is a multipurpose, high-level programming language first released in 1991. It is free, open-sourced and compatible with most operative systems. With a clear and readable syntax, Python is a less error prone software with a large collection of extensions available for different purposes, including statistical, numerical and scientific tools like those found on the SciPy and NumPy extensions. These characteristics make Python an ideal programming language for the development of software packages such as ScrumPy.

### **2.3.2 ScrumPy: metabolic modelling with Python**

ScrumPy (Poolman 2006) is a software package for the reconstruction and analysis of GSMs. It is based on the Python programming language, which is also used as the primary user interface. ScrumPy is a powerful metabolic modelling tool that has been used as the sole modelling tool for several PhD projects and publications (Poolman *et al.* 2009; Poolman *et al.* 2013; Hartman *et al.* 2014) and can be applied for both structural and kinetic modelling. This package interrogates BioCyc (Karp *et al.* 2002) databases and allows null space, elementary modes and linear programming-based analysis of metabolic systems using the Gnu Linear Programming Kit, <http://www.gnu.org/software/glpk/>. ScrumPy is released under the Gnu Public License, and available for download from <http://mudshark.brookes.ac.uk/ScrumPy>. ScrumPy was the software chosen for this project since it provides complete access to the lower-level data structures in the model while effectively hiding them when this is not needed. This leads to a very flexible system, facilitating the production of minor variants on standard analyses for specific purposes while still maintaining the possibility of implementing entirely novel algorithms. ScrumPy also allows examination of the properties of a stoichiometry matrix independently from the rest of the model.

## **2.4 Construction, curation and analysis of the genome-scale model**

This chapter describes the methods utilized for construction, curation and analysis of a genome-scale model of *S. epidermidis*. The GSM was constructed on the basis of a PGDB of the *S. epidermidis* strain RP62A. Using ScrumPy, applying the modelling techniques previously described in Section

2.2 (Poolman *et al.* 2009; Poolman *et al.* 2013; Hartman *et al.* 2014) and using specialised databases and the primary literature, a GSM was generated, curated and analysed for its general properties. Several checks for its general quality assessment were also performed and are described in the following sections.

The criteria considered to choose an appropriate *S. epidermidis* strain for the construction of a genome-scale metabolic model for this project were the following:

- a) the genome annotation of the strain should be fully completed and published.
- b) it should be a clinical strain, involved in foreign body associated infections.
- c) it should be described as a strongly adherent biofilm-forming strain in the primary literature.
- d) it should be available and sequenced in-house.
- e) it should have been screened, together with the rest of the local NAS collection, for biofilm formation and described as a ‘very strong’ biofilm former.
- f) it should be included in one of the main clusters of *S. epidermidis* isolates from the local NAS collection (Section 1.3.2).
- g) it should be potentially transformable.

*S. epidermidis* RP62A (DSM 28319, ATCC 35984) is a biofilm-producing strain isolated in Memphis, Tennessee (United States) during a sepsis outbreak associated with intravascular catheters (1979-1980), whose main pathogenic feature is its ability for cells to accumulate and form biofilms, causing foreign-body infections (Gill *et al.* 2005). This strain, as tested in our laboratory, meets all of the above criteria and was therefore selected for construction of a genome-scale metabolic model of *S. epidermidis*

### 2.4.1 Genome annotation, databases and tools

Annotated genomes and biochemical databases are crucial for the construction and curation of GSMs. A genome annotation assigns specific functions to the sequenced genes of a given organism, while metabolic and biochemical databases generally contain information about the association between genes encoding metabolic enzymes, proteins, enzymes, reactions and metabolites. The construction and curation of the GSM presented on this thesis relied on the extraction and analysis of information from several publicly available databases described here:

BioCyc (Caspi *et al.* 2012), (Karp *et al.* 2017) is a collection of several thousands of organism-specific Pathway/Genome Databases (PGDBs). Single PGDBs are created computationally by the PathoLogic component of the Pathway Tools software (Karp *et al.* 2011; Karp *et al.* 2016) using MetaCyc (Caspi *et al.* 2016). These are collections of database entries such as reactions, metabolites and pathways predicted from the specific genome annotation of each organism plus reactions artificially added by the Pathway Tools software during the ‘gap-filling’ process aimed at minimising the effect of misannotated genes. While this avoids the presence of gaps

in the network, it introduces the risk of overestimating the metabolic potential of the organism and the biological basis of the inclusion of several of these reactions often needs to be verified by consulting alternative databases or the specialised literature. Various versions of the BioCyc PGDB for *Staphylococcus epidermidis* RP62A derived from its annotated genome available at RefSeq were utilised for model construction and curation during the project (Section 2.4.2.2).

The MetaCyc database (Karp *et al.* 2011; Caspi *et al.* 2012) is a PGDB that describes metabolic pathways and enzymes from all domains of life derived from over 46,000 publications. The majority these pathways have been experimentally determined. This is, to date, the largest curated collection of metabolic pathways. MetaCyc was used for manual curation of the GSM.

KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa *et al.* 2000; Kanehisa *et al.* 2002) is a collection of databases useful for understanding high-level functions of a biological system combining genomic and biochemical information. It also includes information at the systems level and regarding the effect of diseases and drugs. KEGG has been found to contain more compound name inconsistencies than BioCyc.

BRENDA (Braunschweig Enzyme Database) (Schomburg *et al.* 2000; Jeske *et al.* 2018) is the main publicly available collection of enzymes and their functionality which data has been directly extracted from literature and critically evaluated by experts. It contains useful information for model curation, including preferred enzyme substrates and cofactors, reaction directionality and reversibility and several kinetic parameters. BRENDA is available at [www.brenda-enzymes.org](http://www.brenda-enzymes.org).

eQuilibrator (Flamholz *et al.* 2012) is a web interface which allows the analysis of reactions under several conditions (e.g. pH, temperature, ionic strength) and provides their corresponding thermodynamic parameters. These can be used to determine, for example, the directionality or reversibility of reactions for which data is either missing or is contradictory in the biochemical databases.

## 2.4.2 Model construction and structure

In an initial step, reactions of the model were extracted from the organism specific PGDB with the ScrumPy module PyoCyc. The model was generated and defined in a modular manner as instructed in Appendix A, Section A.1.1. Splitting certain sets of reactions into different modules allowed for specific analysis of sections with either high metabolic impact (e.g. the electron transport chain) or special interest for the aims of this project (e.g. study of production of biofilm and planktonic biomass components), as well as facilitating the study of the system under different conditions (e.g. utilising different media). The resulting model consisted of seven modules and are described on this section (Figure 2-3):

1. A ‘top-level’ module, which primary function is to import all other modules.
2. Automatically generated reactions (derived from the PGDB)
3. Transporters for media components
4. Transporters for planktonic biomass components
5. Transporters for biofilm biomass components
6. Reactions from the Electron Transport Chain (ETC)
7. Other additional reactions (Extras)

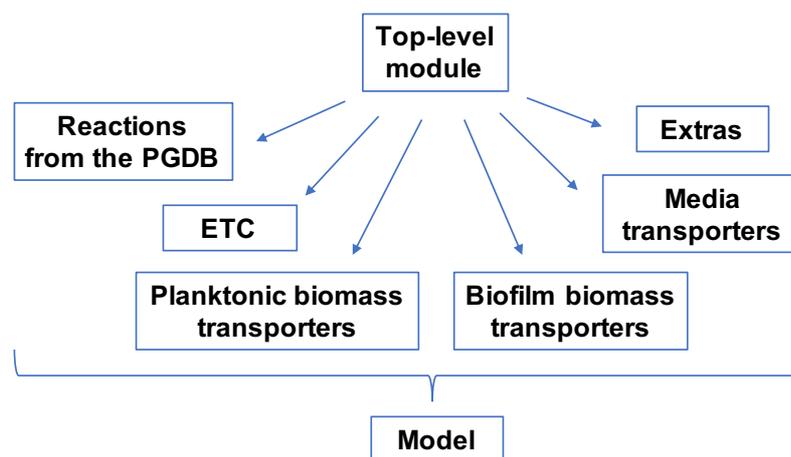


Figure 2-3 Modular structure of the *S. epidermidis* RP62A GSM

#### 2.4.2.1 ‘Top-level’ module

This module specifies that the model is a structural model, the type of data, which metabolites (if any) should be considered as external, and finally, imports the other modules described in multiple files.

#### 2.4.2.2 Automatically generated reactions

The initial reaction set corresponding to ‘*Staphylococcus epidermidis*, strain RP62A’ was extracted from the PGDB flat files obtained from the BioCyc ftp site (Karp *et al.* 2002; Caspi *et al.* 2016). The PGDB used (v. 20.1) had not undergone any manual curation or review (tier 3) and may therefore contain errors, which emphasises the need for subsequent manual curation. Reactions present in this PGDB version were later compared with those in a newer version (v. 20.5) and updated accordingly if needed. However, the most recent version available during the length of this project (v. 22.6) was not used, since it appeared to be less complete: e.g. reactions of the anaerobic respiration seemed to be missing.

#### 2.4.2.3 Transporters

Lack of reliable annotation data for transporters is one of the main challenges in the construction of GSMs. The transporters defined in the PGDB were ambiguous and the data available was far from

complete. Thus, all media and biomass transporters were manually included in the model and described as reactions importing and/or exporting single compounds. Transporting reactions should be updated in the future according to improvements in the databases and more accurate data available at the time.

In the GSM, transport reactions are differentiated from others by addition of the suffix ‘\_tx’ to the reaction name. All external counterparts of internal compounds were differentiated by addition of the prefix ‘x\_’ to the metabolite name. All transporting reactions were defined with the external species on the left hand side, so that a positive flux represents import into the system and a negative flux represents export outside the system.

### **Transporters for planktonic biomass components**

Transporters for individual components of the planktonic biomass were included here. The suffix ‘\_bm\_tx’ was assigned to these reactions for accounting purposes, as described in Poolman *et al.* (2013).

### **Transporters for biofilm biomass components**

Transporters for individual components of the biofilm biomass were added to this module. The suffix ‘\_bf\_bm\_tx’ was assigned to these reactions for accounting purposes.

### **Transporters for media components**

This module contains transport reactions for the import of each compound present in the media. The suffix ‘\_mm\_tx’ was assigned to these reactions for accounting purposes.

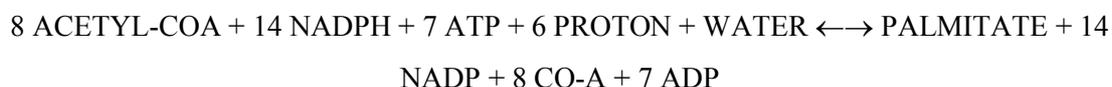
#### ***2.4.2.4 Electron transport chain***

In order to facilitate specific analysis of the electron transport chain (ETC), the ATP synthase reaction and those involved in generation of the proton motive force (PMF) were placed in an independent module. These reactions were initially extracted from the PGDB and modified as follows: most reactions were atomically unbalanced, thus their stoichiometry was corrected as per MetaCyc vs 22.6. Reaction dependency on ubiquinones was substituted for menaquinones, since these are the only quinones known to be present in staphylococci (Sasarman *et al.* 1974; Burke *et al.* 1975; Tynecka *et al.* 1999; McNamara *et al.* 2000).

#### ***2.4.2.5 Additional reactions***

Additional reactions that were required for the individual synthesis of each biomass component were added in a module named ‘Extras’ after confirming that genes for the corresponding enzymes were

present in the genome or the existence of experimental evidence supporting their addition. These included, among others, reactions involved in the synthesis and degradation of ‘glycogen’ (Section 2.4.3.3), and reactions involved in the synthesis of the biofilm polymers PIA1 and PIA2 (Section 2.5). Specific reactions needed for completion of biosynthesis of amino acids were also added, with the basis for their inclusion thoroughly described in Chapter 4 and a generic irreversible ATPase reaction (ATP hydrolysing) was included too. Finally, additional reactions for completion of cell membrane and cell wall synthesis were added: synthesis of cell membrane lipids (palmitate, phosphatidic acid, 1,2-diacylglycerol, CDP-diacylglycerol, phosphatidyl-glycerol and cardiolipin) involves reactions including compounds of undefined atomic composition, which inclusion in the model could complicate identification of potential errors leading to stoichiometry inconsistencies (Section 2.4.3.2). As a way to overcome this problem, a generic lumped reaction synthesising palmitate was included on this module:



This made possible to describe the synthesis of the other lipid-containing compounds as derivatives from palmitate. Note that lumped reactions were represented with lower case to differentiate them from reactions obtained from databases.

The same problem was encountered regarding synthesis of cell wall components (peptidoglycan, wall-teichoic acid and lipoteichoic acid). However, their production is important for biofilm formation in staphylococci (Arciola *et al.* 2012; Becker *et al.* 2014), and so, their full biosynthetic pathways were included, since they will be relevant for future investigations. These pathways are well described on MetaCyc for staphylococci (*S. aureus*) and most of their reactions were already present in the PGDB and therefore in the model. However, the peptidoglycan biosynthetic pathway derived from the PGDB led to production of diaminopimelate-containing peptidoglycan, while it is known that staphylococci produce the lysine-containing type (Somerville 2016). Therefore, some reactions were substituted for their *S. aureus*-specific equivalents on MetaCyc.

### 2.4.3 Model curation

The main aspects of the curation process are explained here in detail. In addition, several quality checks were undertaken in order to ensure consistency between the *in silico* behaviour of the GSM and specific biological features of the modelled organism (e.g. production of experimentally determined biomass components) and are described on the following section. The main steps followed during model curation are summarized in Figure 2-4 and explained below.

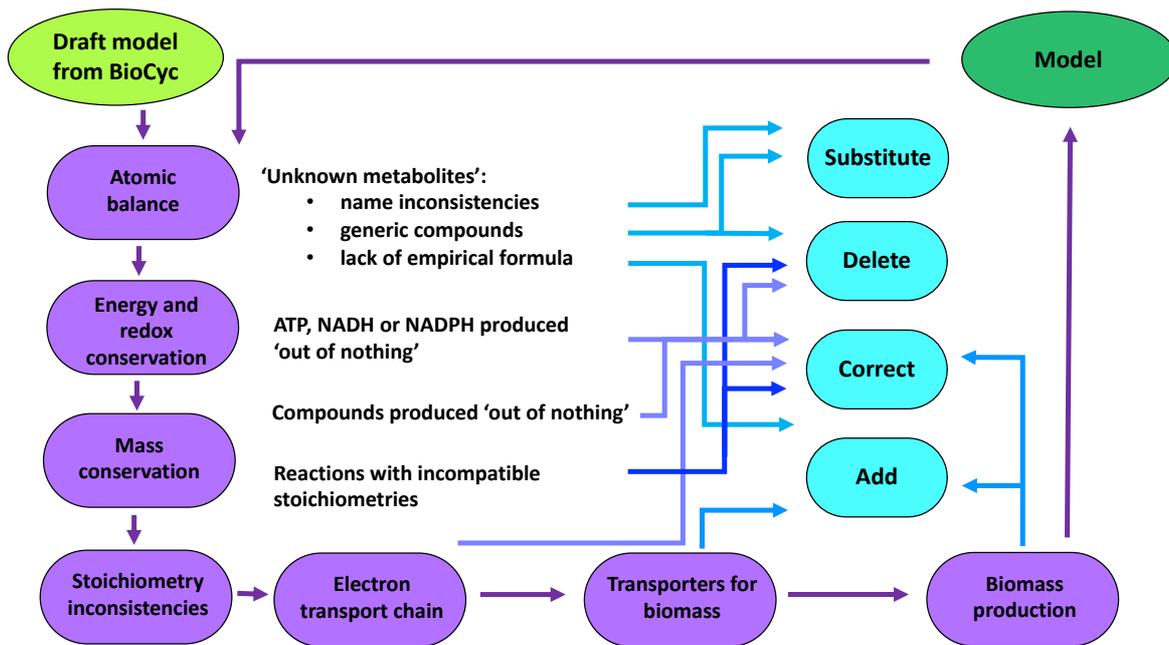


Figure 2-4 Diagram describing the steps that integrate the process of metabolic model construction and refinement.

Adapted from Cell Systems Modelling Group, Oxford Brookes and reproduced here with permission. Purple boxes: aspects or sections of the model that require especial attention during the curation process. Blue boxes: actions undertaken.

### 2.4.3.1 Correction of atomically unbalanced reactions

This step deals with simple material imbalances which are relatively easy to detect. Reactions with wrong stoichiometry derive from errors in the databases used for construction of the model. For example, the reaction RXN-8635 extracted from BioCyc version 20.1 presented the following stoichiometry:



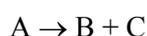
where CPD-8653 is betanidin (with empirical formula  $\text{H}_{14}\text{C}_{18}\text{O}_8\text{N}_2$ ) and CPD-8890 is betanidin quinone (with empirical formula  $\text{H}_{12}\text{C}_{18}\text{O}_8\text{N}_2$ ). This stoichiometry suggests that hydrogen-peroxide ( $\text{H}_2\text{O}_2$ ) dissociates into  $\text{H}_2\text{O}$ , betanidin and betanidin quinone, which is clearly not correct. This inconsistency can be resolved by removing the molecule of betanidin from the right hand side of the equation and adding it to the left hand side, thus allowing hydrogen-peroxide to oxidize betanidin to betanidin quinone while generating two molecules of  $\text{H}_2\text{O}$ .

Sorting out atomic imbalances also implied dealing with 'unknown metabolites': some well-known metabolites were missing assigned empirical formulae and this was simply fixed by manual addition. For those with assigned formulae missing in BioCyc but present in MetaCyc, the information was updated accordingly. For others, relative empirical formulae with respect to related compounds had

to be defined (e.g. reduced *vs* oxidised compound forms). Reactions involved with non-metabolic species (e.g. ‘Damaged-DNA-Pyrimidine’) were removed, while generic unknown compounds such as ‘Amino-Acids’ or ‘Aldehydes’ were either substituted for specific metabolites in stoichiometrically balanced reactions or removed. In this manner ‘Quinones’ and ‘ETC-Quinones’ were substituted for ‘MENAQUINONE’ and ‘D-GLC’ for ‘GLC’ (as per the official metabolite identifier for glucose in ByoCyc), etc. Atomically unbalanced reactions were investigated and corrected by referring to the primary literature and online databases MetaCyc (Caspi *et al.* 2014), BRENDA (Schomburg *et al.* 2000) and KEGG (Kanehisa *et al.* 2002).

### 2.4.3.2 *Material consistency of the model*

Identifying and resolving material inconsistencies caused by the presence of metabolites with unknown empirical formulae is a fairly difficult task. Gevorgyan *et al.* proposed a method capable of detecting these inconsistencies by considering only the stoichiometry of the reactions (Gevorgyan *et al.* 2008). They described an algorithm that, when applied to the analysis of the left null space of the stoichiometry matrix (Section 2.2.1), could identify the metabolites that were unconserved across the system, even when their atomic composition was not known. For instance, given three metabolites (A, B, and C) of atomic masses different to zero, the presence of a set of two reactions such as the ones described below would constitute a stoichiometric inconsistency of the model, since both statements could not be true at the same time:



### 2.4.3.3 *Synthesis of polymers*

In order to avoid introduction of mass inconsistencies, polymer-synthesising reactions were treated in the following manner: polymers were assigned the empirical formulae of their corresponding monomeric units and the stoichiometry of the reactions was adjusted accordingly. For example, two reactions involved in the synthesis and degradation of the GLC polymer ‘glycogen’, initially described in MetaCyc as:

Glycogen synthesis:



Glycogen degradation:



being CPD0-971 a glycogen dextrin, these reactions have net stoichiometries which allow an overall conversion of ADP-D-GLC into ADP, and  $\text{P}_i$  into GLC-1-P, clearly violating the principle of mass conservation:

Glycogen synthesis:



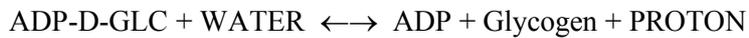
and

Glycogen degradation:



This problem was solved by considering glycogen to be made of one unit of the monomer GLC and re-writing the stoichiometry of the reactions accordingly:

Glycogen synthesis:



Glycogen degradation:



Other polymers specific to this model were also redefined to solve inconsistencies: in the same manner as described for glycogen, reactions involved in the synthesis of the biofilm polymers PIA1 (poly- $\beta$ (1-6)NAcGlc) and PIA2 (deacetylated PIA1) were included in the model. The empirical formula for PIA1 was considered to be that of its monomeric unit NAcGlc, and the reactions were defined as follows:

PIA1 synthesis:



PIA2 synthesis:



A detailed description of the involvement of these PIA polymers in biofilm production, their biochemical composition and their biosynthetic mechanisms is provided in Chapter 0, Section 1.7.4.

#### ***2.4.3.4 Addition of transporters for cell biomass components***

One common way to analyse the production of biomass by a GSM is to define a single lumped biomass equation which specifies the relative proportion of each macromolecular (or micromolecular) biomass component per unit of biomass. This model includes individual

transporters for each single micromolecular biomass component and instead of computing biomass production by defining a biomass equation containing fixed proportions of biomass components, a dictionary of fluxes for these exporters is specified and used as a constraint during LP-based analysis of the system. This allows for a higher degree of detail and flexibility in the study of biomass synthesis since it is possible to, for example, fix the lower flux bounds of these exporters while leaving their upper reaction fluxes unbound. This way, the system is allowed to vary the composition of the biomass produced in response to changing environmental conditions, which is closer to the biological behaviour of the modelled organism.

Due to the lack of detailed biomass composition data available for *S. epidermidis*, transporters for biomass components were added according to the cell composition described in a previous GSM of *S. aureus* (Heinemann *et al.* 2005): this consisted of a representative average biomass composition derived from previously published data for various *S. aureus* strains obtained in several different media and under various conditions. Some modifications were introduced, either for simplicity or accuracy, based upon currently available biochemical data for *S. epidermidis*. In summary, these were: inclusion of free amino acids described in the pool of solutes and wall-associated surface proteins within the protein section; removal of salts from the pool of solutes, since they are not actively involved in any reaction in the model except with their own transport; substitution of GLC equivalents for glycogen as main carbohydrate storage compound; removal of lysyl-phosphatidylglycerol from the lipid section, since it was only present as traces in the cell membrane of *S. epidermidis* (Nahaie *et al.* 1984); and finally, modification of cell wall composition to include the two main polymers in the cell wall of staphylococci (Somerville 2016): a D-D crosslinked peptidoglycan and a peptidoglycan-wall teichoic acid complex, both with complete biosynthetic pathways specifically described for *S. aureus* in MetaCyc. For further details on the specific biomass composition utilised during this project see Appendix B (Section 9.2, Table 9-1).

## 2.4.4 Fundamental validation of the model

After construction, the model was analysed for its properties and several quality checks were performed in order to assure its quality and consistency. The results obtained are presented below, with some specific aspects of the validation process explained in greater detail.

### 2.4.4.1 Model analysis for fundamental validation

Unless stated otherwise, the model was generally analysed by LP as shown in Equation 2-4, selecting minimization of the absolute sum of total fluxes as the objective function and considering this as a proxy for minimising protein investment. The solutions allowed were further constrained by setting the flux through the reactions exporting biomass precursors to meet their proportions experimentally described in the biomass.

The LP was defined as follows (Equation 2-4):

$$\begin{array}{ll}
 \text{minimize:} & \sum_{i=1}^m |v_i| \\
 \text{subject to} & \left\{ \begin{array}{l} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ v_j = t_j ; k \leq j \leq m \\ v_{ATPase} = ATPase \end{array} \right.
 \end{array}$$

where  $v$  is the vector of reaction fluxes and  $N$  is the stoichiometry matrix, with  $n$  rows (metabolites) and  $m$  columns (reactions) and, therefore,  $\mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0}$  defines the steady state assumption.  $v_{k\dots m}$  defines the rates of flux through the biomass transporter reactions from the  $k^{\text{th}}$  to the  $m^{\text{th}}$  reaction of  $N$ , and  $v_{ATPase}$  defines the flux through a generic ATPase reaction, included in the model to meet the energy demand for cell growth and maintenance (e.g. polymerisation of components in the biomass etc).

#### 2.4.4.2 Model-wide conservation of mass

The model was evaluated for mass conservation in the following manner: the flux through all media transporting reactions was constrained to zero, so the model worked as a closed system, while the flux of each exporter for biomass components was set to an arbitrary negative value one at a time. The LP was then solved as described in Equation 2-4. The existence of a feasible solution reflected a violation of mass conservation. In such a case, reactions in the solution were carefully examined, checked against thermodynamic data and corrected or removed as appropriate.

#### 2.4.4.3 Model-wide conservation of energy and redox

The model was evaluated for conservation of energy in the following manner: the flux through all media transporting reactions was constrained to zero, so the model worked as a closed system, while the flux through the generic ATPase reaction was set to an arbitrary positive value. The LP was then solved: the existence of a feasible solution reflected a violation of energy conservation. In such a case, reactions in the solution were carefully examined, checked against thermodynamic data and corrected or removed as appropriate. A similar procedure was used to ensure redox conservation by detecting oxidation of the reducing equivalents NADH and NADPH.

#### 2.4.4.4 Feasibility of biomass production

The capability of the model to produce each single biomass component at a time was investigated: if the LP had no feasible solution, two possibilities were considered: i) missing reactions could be causing a discontinuity in the network, thus preventing biosynthesis of the compound, or ii) all reactions needed for its synthesis were present but some could exhibit wrong directionality or reversibility. In the first case, missing reactions were added to the model and in the latter, candidate reactions for thermodynamic re-definition were identified in the following manner: all reactions in the model were temporarily declared reversible and the LP problem was solved again. If a feasible

solution was found, reactions previously defined as irreversible, which flux sign had changed in the new solution, therefore, working in the opposite direction were evaluated against thermodynamic data and their reversibility or directionality was then re-defined as corresponded. Energy and redox conservation properties were re-checked after each alteration.

#### **2.4.4.5 Definition of auxotrophies and essential media components**

The LP-based analysis technique was used to check for the essentiality of media components. This allowed defining possible auxotrophies to be compared with experimental data for model curation and to detect possible gaps in the network due to missing reactions that needed to be included: initial investigations considering an *in silico* Glc-based minimal medium with  $\text{NH}_4^+$  as sole N source demonstrated that the model was not able to account for the synthesis of several biomass components, including the amino acids Asn, Cys, Met and Phe. Since staphylococci have been reported to present a wide range of amino acid, purine and vitamin auxotrophies, both for growth and biofilm production (Knight 1937; Emmett *et al.* 1975), it was difficult to discern if these auxotrophies represented real biological features of the organism or reflected errors or gaps in the system. Therefore, experimental data on minimal growth requirements was obtained and used for further model refinement (Chapter 4). After this, the model was again analyzed to check for the essentiality of media components: it could finally reproduce synthesis of each individual biomass component utilising an *in silico* minimal medium composed of Glc, core set substrates ( $\text{NH}_4^+$ ,  $\text{SO}_4^{2-}$  and  $\text{HPO}_4^{2-}$ ),  $\text{H}_2\text{O}$  and solely supplemented with the vitamin niacin. If niacin is not provided, the system cannot account for the synthesis of NAD, NADH, NADP or NADPH.

#### **2.4.4.6 Other quality checks**

Inconsistent reaction subsets define reactions which cannot carry flux at the same time in any steady state solution. They can represent potential errors in the description of the model due to missing reactions or reactions defined inappropriately. Determining the inconsistent reaction subsets in the model allowed identification of discontinuities in the network (i.e. caused by metabolite name inconsistencies) and reactions with inappropriate directionality or reversibility. These inconsistencies were resolved by addition of missing reactions or by modifying existing reactions upon evaluation of information from the databases (MetaCyc, BRENDA and eQuilibrator (Flamholz *et al.* 2012)) and the primary literature.

## **2.5 General properties of the model**

The general model properties were examined by null space and LP-based analysis of the stoichiometry matrix and are summarized in Table 2-1:

Table 2-1 General properties of the GSM for *S. epidermidis* RP62A

<b>FEATURES OF THE GSM</b>	<b>BEFORE CURATION</b>	<b>AFTER CURATION</b>
<b>Reactions excluding transporters</b>	1264	880
<b>Transporters</b>	42	72
<b>Reactions associated with identifiable genes</b>	Not computed	606
<b>Metabolites</b>	1411	859
<b>Dead reactions</b>	767	339
<b>Orphan metabolites</b>	624	307
<b>Metabolites with undefined empirical formulae</b>	564	0
<b>Unbalanced reactions</b>	579	0
<b>Atomic balance for C, N, S, P, O, H</b>	No	Yes
<b>Stoichiometry inconsistencies</b>	Yes	No
<b>Reaction subsets</b>	388	406
<b>Inconsistent reaction subsets</b>	29	0
<b>Mass, energy and redox conservation</b>	No	Yes
<b>Biomass production</b>	23 out of 49 components	All components except niacin (auxotrophy)

This table summarizes how the properties of model have been modified throughout the curation process and compares the characteristic of the draft model with its final curated version.

Iterative refinement of the initial draft model automatically derived from the PGDB of *S. epidermidis* RP62A into the final GSM presented here implied removal of 384 reactions and 552 metabolites in order to deal with problems such as presence of reactions wrongly included in the PGDB for this organism, solve inconsistencies derived from the use of multiple names to refer to single compounds, or deal with reactions with incompatible stoichiometries, as already described in this chapter. Over one third of the reactions of the final model are dead and over a third of the metabolites present are identified as 'orphan'. This is not an unusual result and it is mainly due to the genetic and metabolic information available for this organism being incomplete.

Several quality checks were performed in order to assure the quality and coherence of the GSM: after curation, all reactions were atomically balanced for those elements accounting for the major atomic composition of the biomass: carbon (C), nitrogen (N), phosphorous (P), sulphur (S), oxygen (O) and hydrogen (H) (Section 2.4.3.1). An artificial metabolite (x\_Awork) was introduced as a product of the ATPase reaction for accounting purposes and has no mass and no associated empirical formulae. Therefore, an undetermined atomic imbalance was introduced related to this compound but has no real significance. For the same reasons, this is also the sole metabolite for which overall stoichiometry is not conserved throughout the network but this does not reflect a real error in the model description (Section 2.4.3.2). Furthermore, the model presented several mass, energy and redox inconsistencies, all of which were eliminated during the curation process following the methods described in Sections 2.4.4.2 and 2.4.4.3. Finally, a total of 29 inconsistent subsets were determined, investigated and resolved (Section 2.4.4.6).

In summary, a GSM of *S. epidermidis* RP62A has been constructed, curated and analysed for its fundamental properties. The model is conserved for mass, energy and redox and free of stoichiometric inconsistencies and can account for production of each individual biomass component utilising an *in silico* Glc-based minimal medium supplemented with niacin. This chapter described the construction, curation and validation processes of the GSM and established its good quality, hence supporting its use as a tool to study the metabolism of *S. epidermidis* RP62A.

Relevant code to the work described in this chapter can be found in Appendix A (Sections 9.1.2.1, 9.1.2.2, 9.1.2.4 and 9.1.2.7);

## 3 Fundamental characterisation of the genome-scale model

### 3.1 Introduction

Bacteria are capable of adjusting their metabolism to maintain required ATP yields upon changes in environmental conditions such as variation in the availability of nutrients or electron acceptors (Uribe-Alvarez *et al.* 2016), therefore, understanding the mechanisms behind these processes gives useful insight into their ability to adapt to different hosts and environments (e.g. growing on the skin vs growing in joints in the human body).

The previous chapter described the general properties of the model constructed for *S. epidermidis* RP62A and how several quality checks applied demonstrated its consistency, hence supporting its application to the study of the bacterial metabolism. Here we describe how the model was used to investigate biologically relevant metabolic states. The analysis undertaken allowed comparison of the model's behaviour with biological features and experimentally described strategies exerted by staphylococci, such as the functioning of the electron transport chain and utilisation of various C sources for energy production. Specifically, the model was utilised to account for: i) Glc, Glt and Ac utilisation for ATP synthesis; ii) production of planktonic biomass; and iii) production of acetoin and butanediol as well as their utilisation as C sources, as has been described on staphylococcal biofilms. During these analysis, the system was used to explore likely responses to changes in the availability of electron acceptors ( $O_2$  and  $NO_3^-$ ).

The relevant background for these investigations is described in Chapter 0, Section 1.4.

### 3.2 Methods

#### 3.2.1 Model analysis for ATP production

LP-based analysis of the model was applied to study the capacity of the system to generate ATP from various C sources without considering biomass production, thus allowing definition of the system's maximum theoretical ATP yields. The analysis assumed that the *in silico* medium available to the system was a minimal medium composed of  $SO_4^{2-}$ ,  $HPO_4^{2-}$ ,  $NH_4^+$ ,  $H_2O$ , niacin and the C source/s relevant to each particular study, plus additional  $NO_3^-$  in order to enable anaerobic respiration. The model utilised for this work was Sepi\_MinMed.spy unless stated otherwise. The LP formulation applied was described as follows (Equation 3-1):

$$\text{i) minimize: } \sum_{i=1}^m |\mathbf{v}|$$

or

$$\text{ii) minimize: } \mathbf{V}_{\text{Glc\_tx}}$$

$$\text{subject to } \left\{ \begin{array}{l} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ \mathbf{v}_{\text{ATPase}} = \text{ATPase} = 1 \end{array} \right.$$

where the flux through the ATPase reaction was set to 1, forcing the system to consume 1 unit of ATP. While this value does not have a physiological significance it was chosen as the ATPase flux constraint in all analysis for production of ATP in this chapter in order to facilitate the comparison of the results obtained. The LP was solved twice, selecting as objective function either: i) minimization of the sum of total fluxes through the network as a proxy for minimizing the enzymatic cost (based on the speculation that a higher flux would be equivalent to a higher enzymatic activity); or ii) minimization of flux through the C source importer, thus optimising its utilisation for ATP production, since by reducing the consumption rate of these substrates to the minimum while maintaining the ATP demand constant the solution obtained will produce the higher ATP per unit of C source yield. The analysis was performed in aerobic and anaerobic conditions and in the presence and absence of  $\text{NO}_3^-$ .

Relevant code to the work described in this chapter can be found in Appendix A, Sections 9.1.2.1, 9.1.2.3, 9.1.2.6, 9.1.3.1 and 9.1.3.2)

## 3.3 Results

### 3.3.1 Characterisation of the electron transport chain

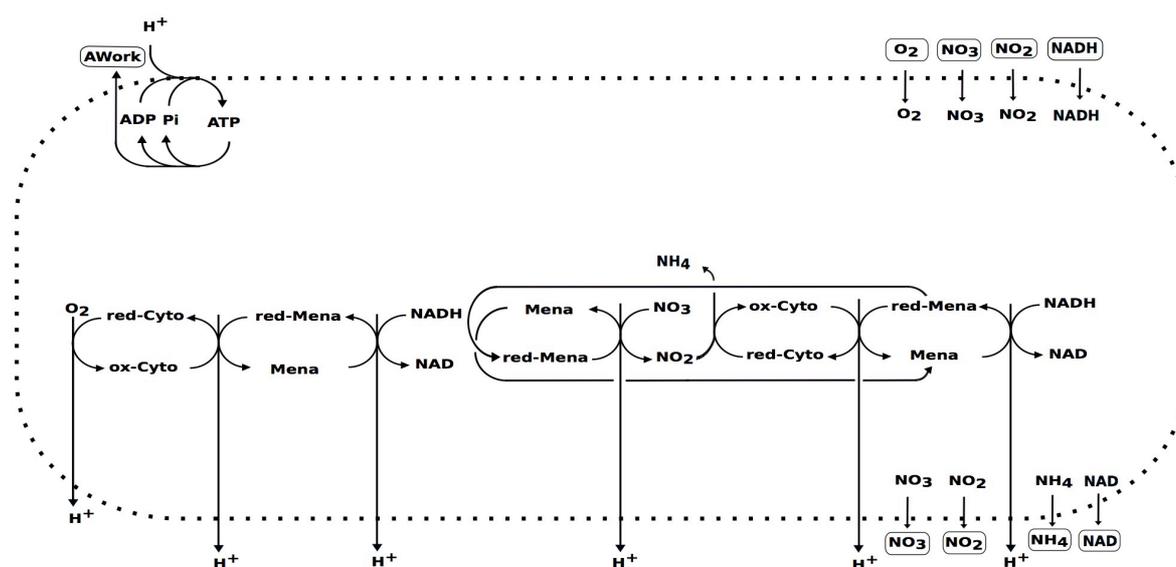
#### Model analysis

Reactions involved in the ETC were re-written as a stand-alone module, which included all reactions present in the ETC module described in Chapter 2, Section 2.4.2, plus those allowing regeneration of oxidized cytochromes via reduction of either  $\text{NO}_3^-$  or  $\text{NO}_2^-$  during anaerobic respiration. This allows analysis of the ETC as an independent system, which was needed in order to ensure its correct functioning as part of the model curation process. Its analysis identified all independent minimal routes for ATP synthesis through this subnetwork by calculating its elementary modes (EMs). This allowed comparison of the ATP production efficiency of these routes by calculating their net stoichiometries and their ATP/NADH ratios. An artificial metabolite ( $x_{\text{Awork}}$ ) was introduced in the stand-alone module as a product of the ATPase reaction. This was purely done for accounting purposes, since it allowed identifying the flux through the ATPase reaction from the net

stoichiometry of the calculated EMs. Thus production of one unit of  $x\_Awork$  would denote one unit of flux through the ATPase reaction (sole ATP consuming reaction in this system), and therefore, 1 unit of flux through the ATP synthase reaction. Since  $x\_Awork$  has no mass or associated empirical formulae, stoichiometric imbalances related to this compound can be considered trivial. For completion, the  $NO_3^-$  and  $NO_2^-$  transporters of the stand-alone ETC were defined as reversible, although were made irreversible during further analysis of the model, with  $NO_3^-$  being imported and  $NO_2^-$  exported as corresponded, according to the specific conditions considered in each study.

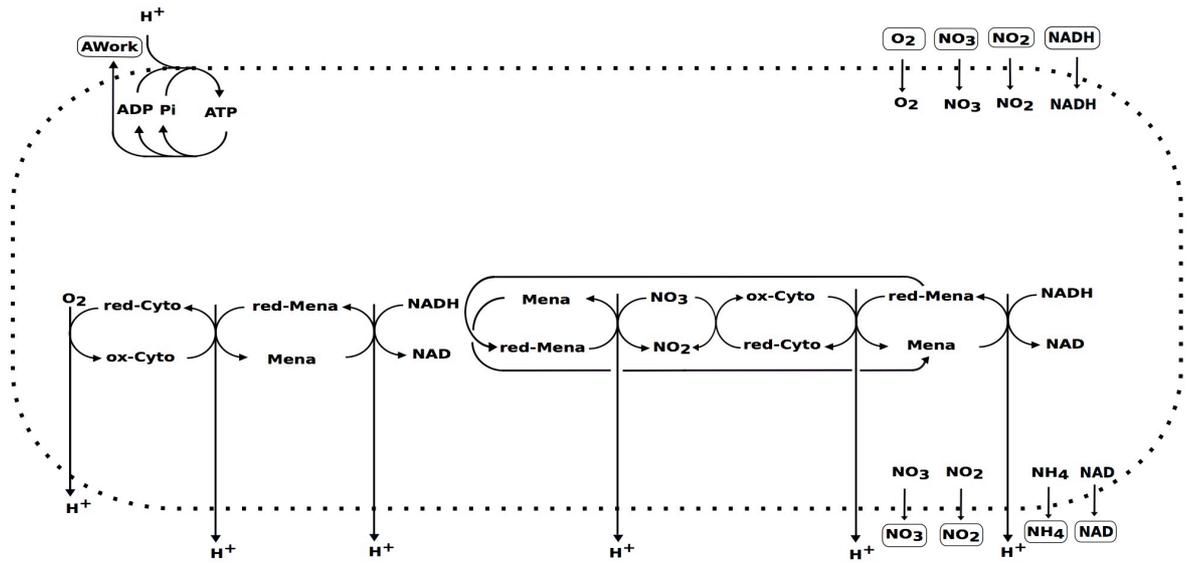
## Results

The figures shown below describe the reactions contained in this module and their potential interaction patterns: while Figure 3-1 represents the menaquinone-dependent  $NO_3^-$  reduction mechanism described for staphylococci (Chapter 0, Section 1.4.1), Figure 3-2 represents its cytochrome-dependent alternative (Chapter 0, Section 1.4.1). For simplicity,  $H_2O$  and internal protons are not shown in these diagrams.



**Figure 3-1 Reactions of the ETC representing the menaquinone-dependent  $NO_3^-$  reduction mechanism described for staphylococci.**

The abbreviations shown in this figure stand for the following: AWork = artificial external metabolite introduced as a product of the ATPase reaction for accounting purposes with no mass or associated empirical formulae; Mena = menaquinone; Cyto = cytochrome; red = reduced; ox = oxidised. The  $NO_3^-$  reduction mechanism represented here is functionally linked to the oxidation of menaquinones and can work without directly involving the action of cytochromes (Sasarman *et al.* 1974; Uribe-Alvarez *et al.* 2016).



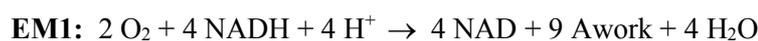
**Figure 3-2 Reactions of the ETC alternative mechanism of  $\text{NO}_3^-$  reduction via cytochromes described for staphylococci**

The abbreviations shown in this figure stand for the following: AWork = artificial external metabolite introduced as a product of the ATPase reaction for accounting purposes with no mass or associated empirical formulae; Mena = menaquinone; Cyto = cytochrome; red = reduced; ox = oxidised. Here,  $\text{NO}_3^-$  is directly reduced upon interaction with cytochromes which get oxidised and are brought back to their reduced state by oxidising menaquinones (Sasarman *et al.* 1974; Burke *et al.* 1975; Heinemann *et al.* 2005).

If the ETC follows the menaquinone-dependent  $\text{NO}_3^-$  reduction mechanism, the  $\text{NO}_2^-$  produced could be either reduced further to  $\text{NH}_4^+$  via interaction with cytochromes or be excreted out of the system. However, the cytochrome-dependent mechanism does not allow the resulting  $\text{NO}_2^-$  to be reduced further, hence its excretion becomes mandatory.

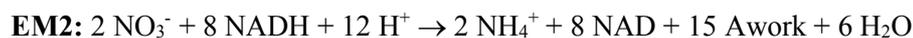
A total of 6 EMs were determined during the analysis, presenting 4 different net stoichiometries. Their net stoichiometries were calculated and are described below. The ATP/NADH and P/O ratios of these modes were calculated in the following manner: for example, in EM1 (below), 4 molecules of NADH were oxidized to NAD and 9 molecules of ATP were produced (denoted by Awork). Thus, its associated ATP/NADH ratio would correspond to  $9 \text{ ATP} / 4 \text{ NADH} = 2.25$ . Since 2 molecules of  $\text{O}_2$  are consumed in the process and this is equivalent to 4 half molecules of  $\text{O}_2$  (4 O), the P/O ratio for EM1 would be  $9 \text{ ATP} / 4 \text{ O} = 2.25$ .

### Net stoichiometry and ratios of modes utilising O<sub>2</sub> as final electron acceptor



ATP/NADH ratio = 2.25; P/O ratio = 2.25

### Net stoichiometry and ratios of modes utilising NO<sub>3</sub><sup>-</sup> as final electron acceptor

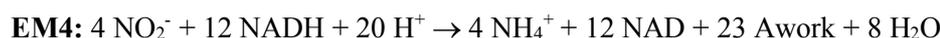


ATP/NADH ratio = 1.87



ATP/NADH ratio = 1.75

### Net stoichiometry and ratios of modes utilising NO<sub>2</sub><sup>-</sup> as final electron acceptor



ATP/NADH ratio = 1.92

The excretion of protons as H<sub>2</sub>O in all EMs is compensated by the import of external protons via the ATP synthase reaction.

## 3.3.2 Metabolic responses for ATP production under a range of environmental conditions

### 3.3.2.1 ATP production from glucose

In order to investigate the ability of the system to obtain energy from Glc and reproduce the behaviour described for staphylococci during the first metabolic state (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016), an *in silico* study was performed where the system was analysed for ATP production utilising a Glc-based minimal medium.

#### Model analysis

The model was analysed as described in Section 3.3.2, considering Glc as the sole C source available for ATP synthesis and in the presence and the absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>.

#### Results

The optimal solutions obtained for the conditions tested are compared in Table 3-1:

**Table 3-1 Model responses for production of ATP from Glc considering two alternative analysis objectives in the presence and absence of electron acceptors**

Analysis objective	Minimization of total net flux through the network				Minimization of Glc import			
Electron acceptor available	ATP per C	ATP per Glc	P/O ratios	Metabolic strategy followed	ATP per C	ATP per Glc	P/O ratios	Metabolic strategy followed
<b>O<sub>2</sub></b>	2.00	12.00	2.00 3.00	Oxidation to Ac + ETC with O <sub>2</sub>	4.33	26.00	1.83 2.17	Glycolysis + TCA cycle + ETC with O <sub>2</sub>
<b>NO<sub>3</sub><sup>-</sup></b>	1.75	10.5	-	Oxidation to Ac + ETC with NO <sub>3</sub> <sup>-</sup>	3.58	21.5	-	Glycolysis + TCA cycle + ETC with NO <sub>3</sub> <sup>-</sup>
<b>None</b>	0.667	4.00	-	Fermentation to butanoate and Form	0.667	4.00	-	Fermentation to butanoate and Form

Conditions set: presence of O<sub>2</sub>, anaerobiosis and presence and absence of NO<sub>3</sub><sup>-</sup>. When both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> are available, O<sub>2</sub> was the only electron acceptor utilised, thus this condition was omitted from the table. For those conditions showing two P/O ratio values, the first value corresponds to the oxidative plus substrate level phosphorylation and the second one to the oxidative phosphorylation only.

These LP analysis solutions are represented in diagrams and described in further detail below (Figure 3-3 to Figure 3-6).

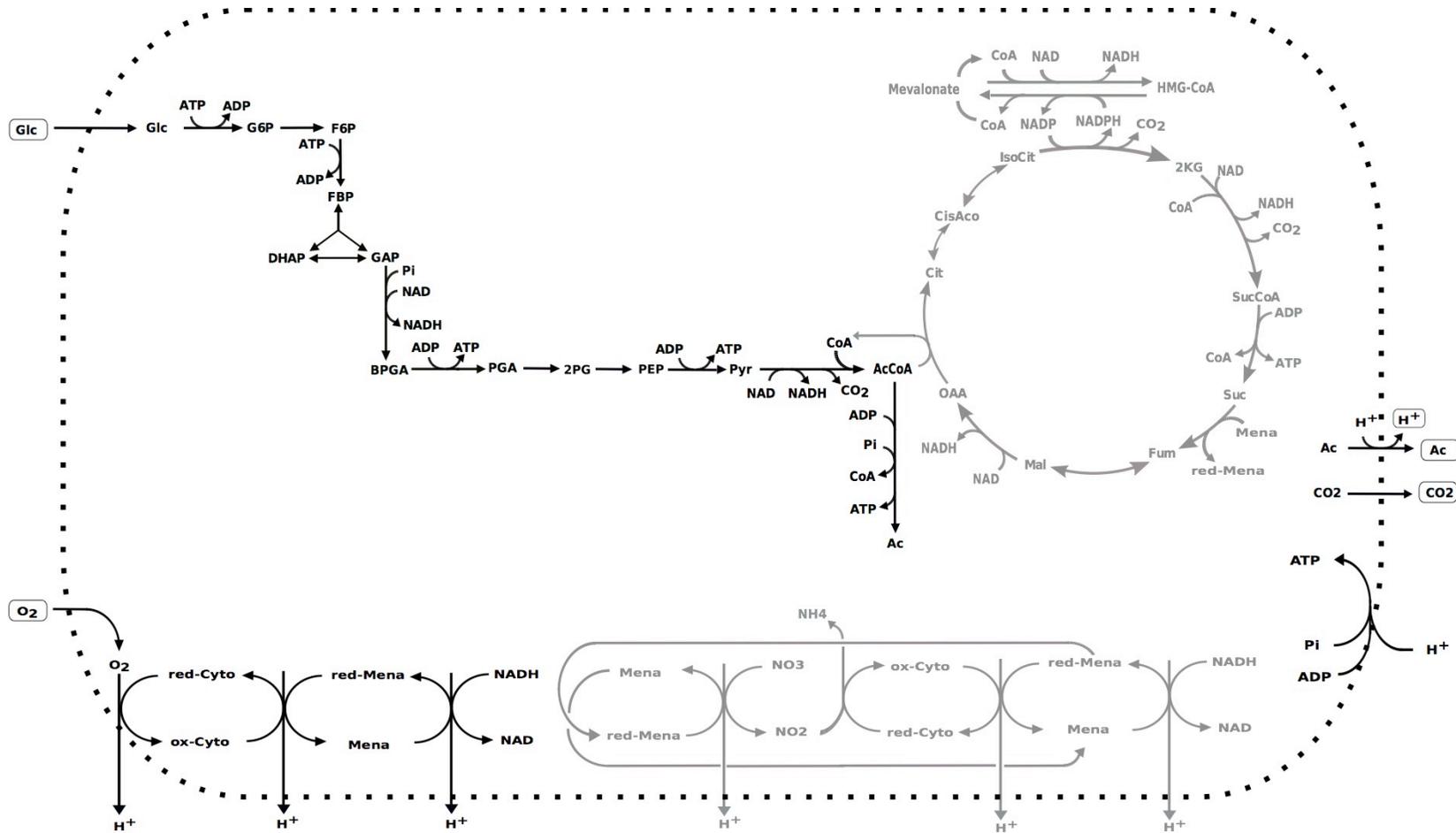


Figure 3-3 ATP production from Glc in the presence of O<sub>2</sub> when total flux through the system was minimised.

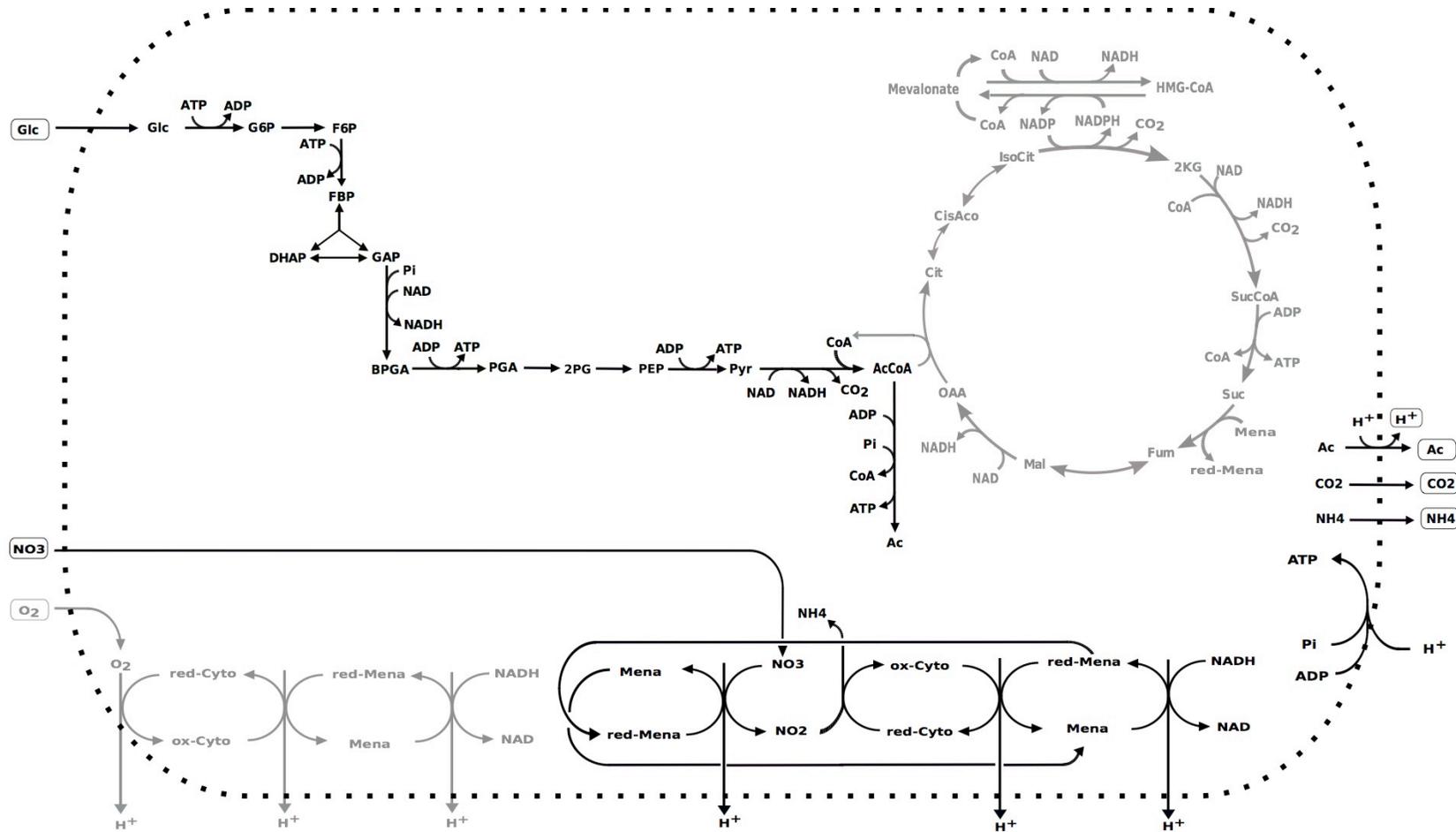


Figure 3-4 ATP production from Glc in the presence of NO<sub>3</sub><sup>-</sup> when total flux through the system was minimised.

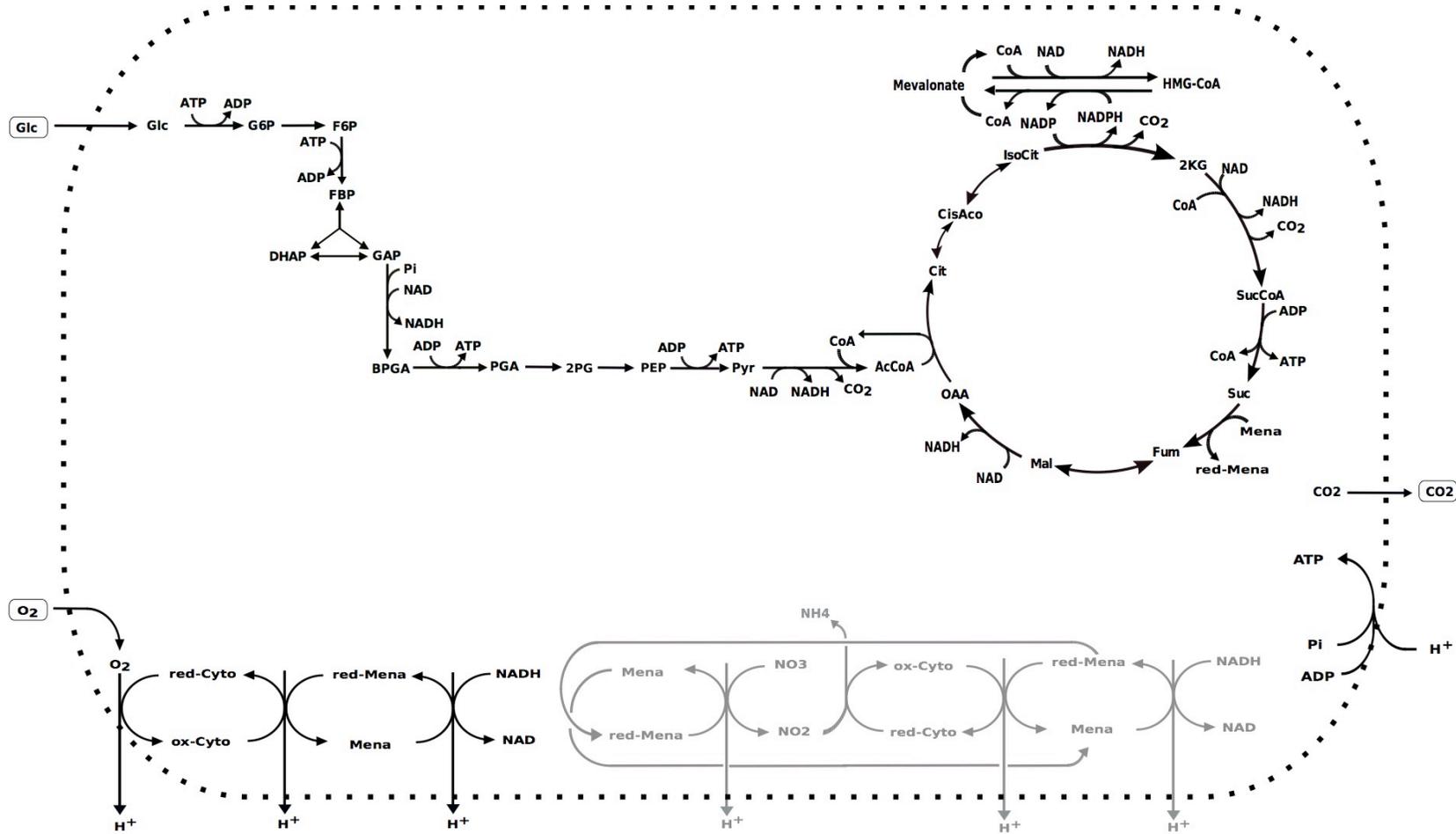


Figure 3-5 ATP production from Glc in the presence of  $O_2$  and  $NO_3^-$  when the import of Glc was minimised.

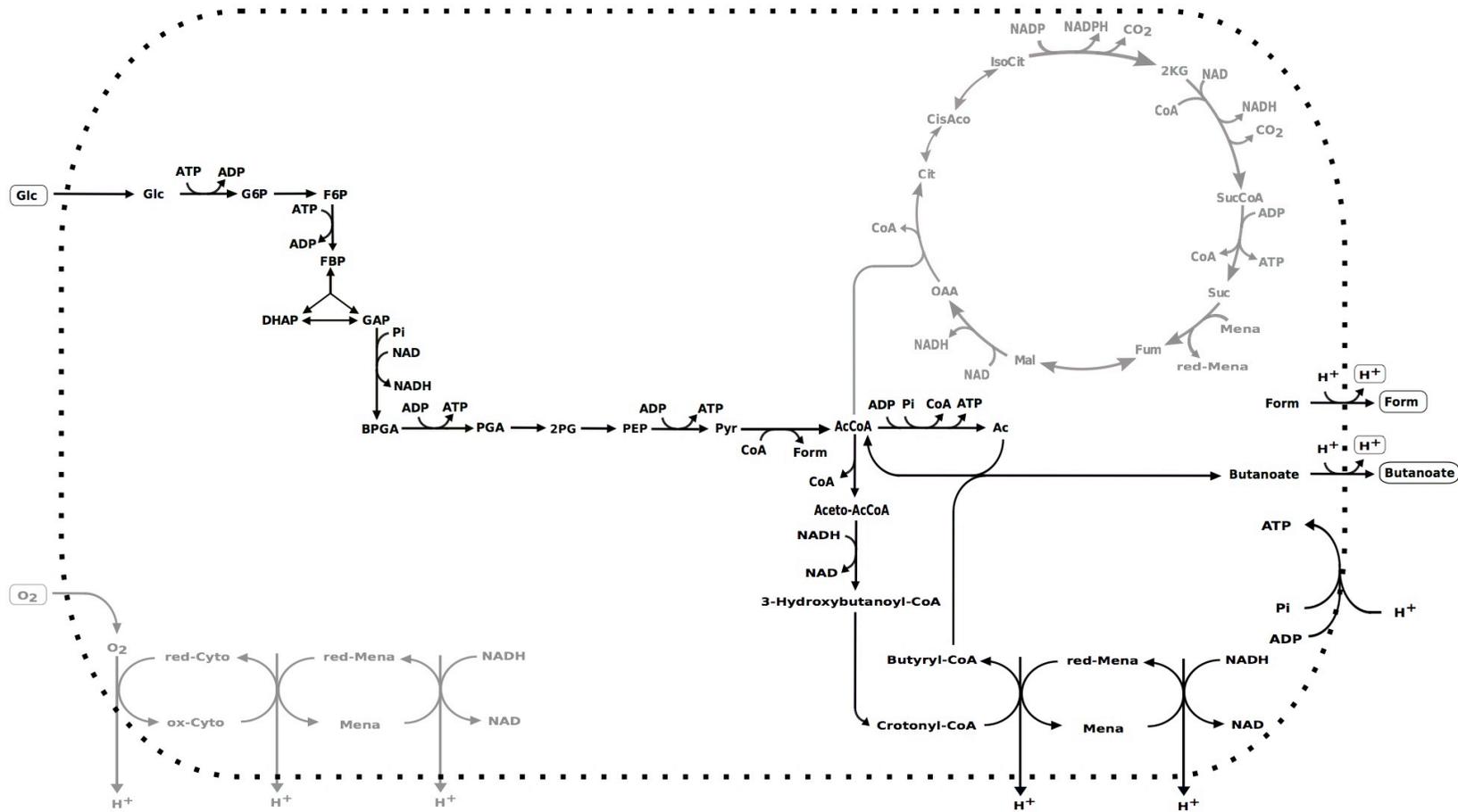


Figure 3-6 ATP production in the presence of Glc and the absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> for both objective functions: minimisation of net total flux through and minimisation of Glc consumption)

In the response obtained for ATP production from Glc in the presence of O<sub>2</sub> when total flux through the system was minimised (Figure 3-3), Glc underwent glycolysis and the Pyr produced was oxidized to Ac via AcCoA by Ac-CoA ligase (EC 6.2.1.1) working in the reverse direction, rather than by the Pta-AckA pathway commonly described for staphylococci (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016; Halsey *et al.* 2017) (Chapter 0, Section 1.4.2.1). Since the Ac-CoA reaction presents the same net stoichiometry as this pathway, it was prioritised by the system in order to reduce the total net flux through the network. The NADH generated entered the ETC, where electrons were finally transferred to O<sub>2</sub> and NAD was regenerated to maintain the glycolytic flux and redox balance. 12 molecules of ATP were produced per molecule of Glc consumed. The protons excreted with Ac contributed to the PMF that pumps the ATP synthase.

When O<sub>2</sub> was absent but NO<sub>3</sub><sup>-</sup> was available, a similar response was observed (Figure 3-4), however, in this case NO<sub>3</sub><sup>-</sup> was utilised as the ultimate electron acceptor of the ETC, generating 10.5 molecules of ATP per Glc consumed. Here, the protons excreted with Ac (but not NH<sub>4</sub><sup>+</sup>) contributed to generate the PMF that enabled flux through the ATP synthase. This solution matches experimental observations published by Uribe-Alvarez *et al.* 2016, where the menaquinone-dependent NO<sub>3</sub><sup>-</sup>-reducing mechanism was seen to be favoured *in vitro* over its cytochrome-NO<sub>3</sub><sup>-</sup> reducing alternative, as previously described in Section 1.4.1

When metabolism was re-directed to optimise Glc utilisation for energy production (Figure 3-5), the model catabolised Glc via glycolysis and the TCA cycle. O<sub>2</sub> was utilised as the final electron acceptor of the ETC, with the subsequent generation of 26 ATP molecules per molecule of Glc. When O<sub>2</sub> was not available, a similar response was observed (data not shown), utilising NO<sub>3</sub><sup>-</sup> as the ultimate electron acceptor and generating 21.5 molecules of ATP per Glc consumed. Since the TCA cycle of prokaryotes works with a NADPH-dependent isocitrate-dh (MetaCyc vs 22.6), the action of a transhydrogenase becomes mandatory in order to allow the NADPH produced to be ‘translated’ into NADH, as was shown in this response. The NADH produced entered the ETC for ATP production and regeneration of NAD for the glycolytic process.

When neither O<sub>2</sub> or NO<sub>3</sub><sup>-</sup> were available (Figure 3-6), Glc was fermented to Form, Ac and butanoate. Ac was not excreted to the medium but rather utilised to regenerate AcCoA during reduction of butyryl-CoA to butanoate. This strategy was less efficient than previous responses, with only 4 molecules of ATP being produced per Glc consumed.

When a further constraint was applied and the flux through the Form exporter was set to zero, the model fermented all Glc to Lac in a less efficient process, generating only 2 molecules of ATP. This proves that even though Glc fermentation to Lac is possible (as reported in the literature (Sivakanesan *et al.* 1980; Fuchs *et al.* 2007)), it is less efficient than other alternative fermentative

responses, not only from an energetic point of view, but also with respect to the total net flux through the network, since the objective value of this response is higher than that obtained during fermentation to butanoate and Form. Protons excreted with butanoate and Form did not contribute to the PMF needed to pump the ATP synthase. Note that, while the LP-based analysis solution obtained for minimisation of Glc consumption was similar to the one obtained for minimisation of total net flux through the network, in this case the action of the Pyr kinase was substituted by two alternative reactions with the same net stoichiometry, yielding a solution with a slightly higher total net flux (+ 0.5 mmol/gDW/h, where DW indicates dry weight) but the same Glc consumption rate.

Although Lac and Ac excretion was not observed in the absence of electron acceptors (Figure 3-6), this does not rule out the possibility that it would occur in other optimal or sub-optimal solutions. In fact, 4 out of the 7 reactions involved with the pathway of Pyr fermentation to butanoate are catalysed by enzymes with no associated genes and, therefore, have been included in the database via automatic gap-filling. Blocking flux through the short chain acyl-CoA-dh (EC 1.3.8.1) reaction, with no associated gene, prevented fermentation to butanoate or butanol and led to Form, Ac and EtHO production and excretion ( Table 3-2 and Figure 3-7).

**Table 3-2 Model response for production of ATP from Glc considering two alternative analysis objectives in the absence of electron acceptors**

Analysis objective	Minimization of total flux or minimization of Glc import			
	ATP per C	ATP per Glc	P/O ratios	By-products exported
None	0.5	3.0	-	Ac, EtHO, Form

Conditions set: absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> and blocking of flux through the short chain acyl-CoA-dh reaction. The responses obtained considering either of the alternative analysis objectives presented the same main characteristics.

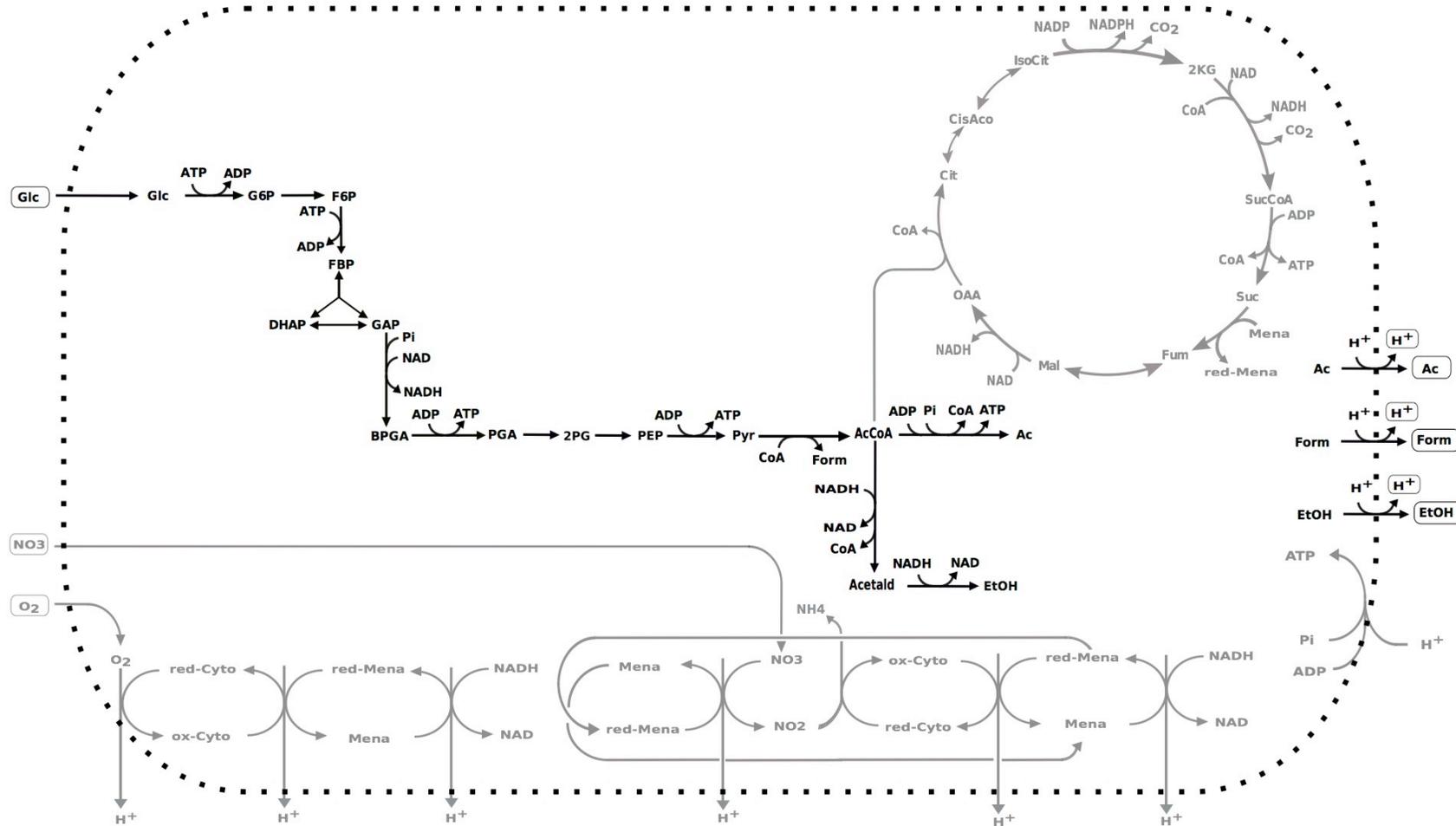


Figure 3-7 ATP production in the presence of Glc and the absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> when total flux through the system was minimised and flux through the reaction catalysed by the short chain acyl-CoA-dh (EC 1.3.8.1) was blocked.

Blocking flux through the short chain acyl-CoA-dh reaction led to Form, Ac and EtHO production and excretion (Figure 3-7). ATP was generated via glycolysis and Pyr fermentation to Ac and the redox balance was maintained via NADH oxidation to NAD during Pyr fermentation to EtHO. This optimal solution did not involve the activity of the ATP synthase. Although less efficient than the previous response (Figure 3-6), since 3 ATP molecules were generated instead of 4, this one is potentially more consistent with the observed biological behaviour of staphylococci (Sivakanesan *et al.* 1980; Fuchs *et al.* 2007) but the fact that the living organism could follow different metabolic strategies (including the one described here) upon variation in the experimental conditions cannot be ruled out. Therefore, no alterations to the system were introduced at this stage and the current version of the model includes this reaction. This should be considered when developing future work, and the model modified accordingly if necessary.

### 3.3.2.2 ATP production from glutamate

The only experimental work describing a specific P/O ratio for staphylococci gave a value of 1.5 for both substrate level and oxidative phosphorylation, and was obtained from a nutrient-depleted *S. aureus* culture supplemented with Glt as sole C source (Tynecka *et al.* 1999; Heinemann *et al.* 2005). In order to compare the system's behaviour with these *in vitro* data, the model was analysed for production of ATP from Glt.

#### Model analysis

For this purpose, the model was analysed for ATP production as described in Section 3.3.2, with Glt substituting Glc as the sole C source present in the *in silico* minimal medium.

#### Results

These LP analysis solutions are represented in diagrams and described in further detail in Figure 3-8 to Figure 3-9.

The P/O ratios associated with these responses are summarised in Table 3-3:

**Table 3-3 P/O ratios calculated for the LP analysis solutions obtained for ATP production from Glt considering two alternative analysis objectives in the presence of O<sub>2</sub>**

P/O ratios	Minimization of total flux	Minimization of Glt import
Oxidative phosphorylation	2.25	1.83
Oxidative plus substrate level phosphorylation	2.75	2.06

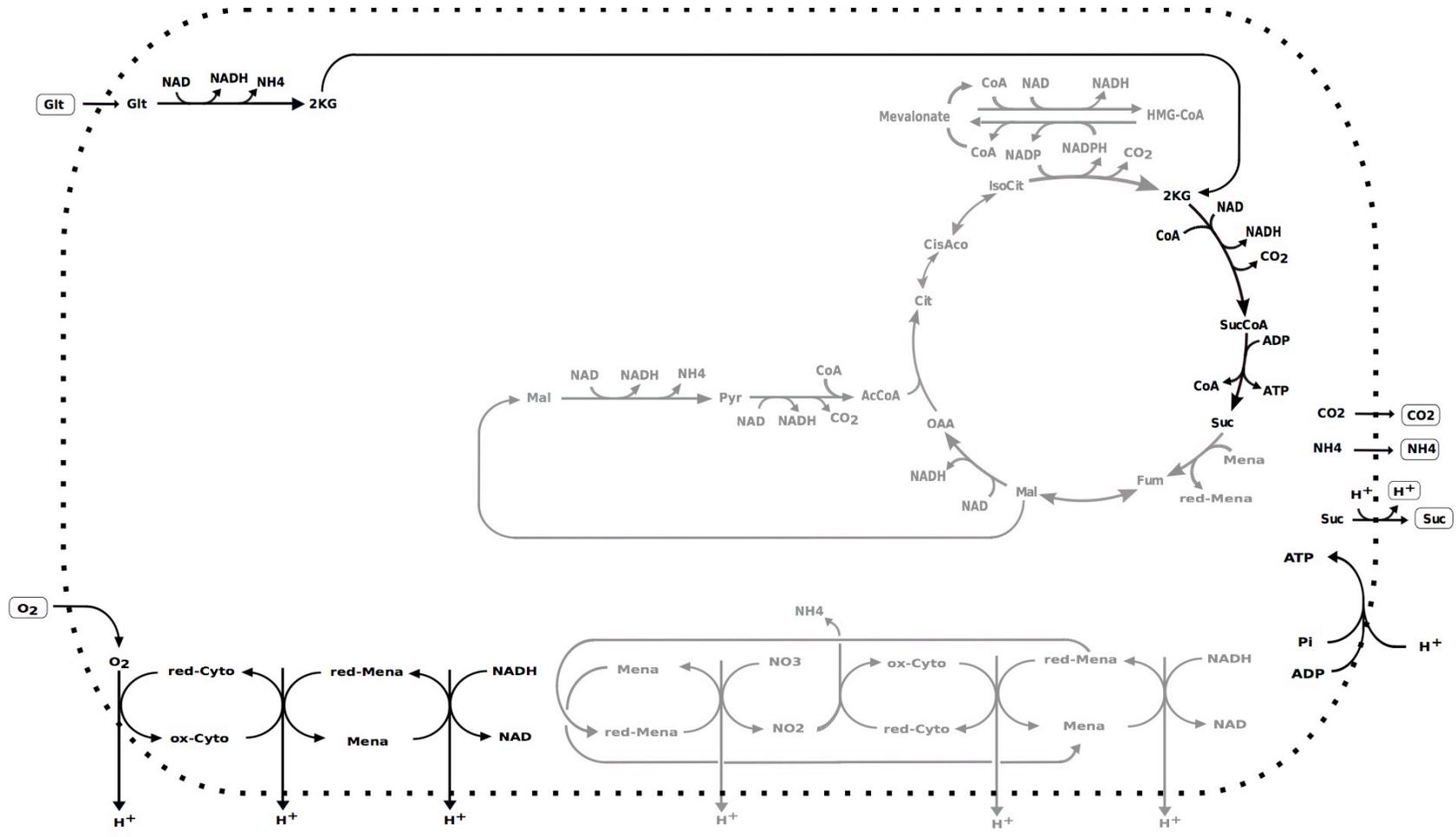


Figure 3-8 ATP production from Glt in the presence of O<sub>2</sub> when total flux through the system was minimised.

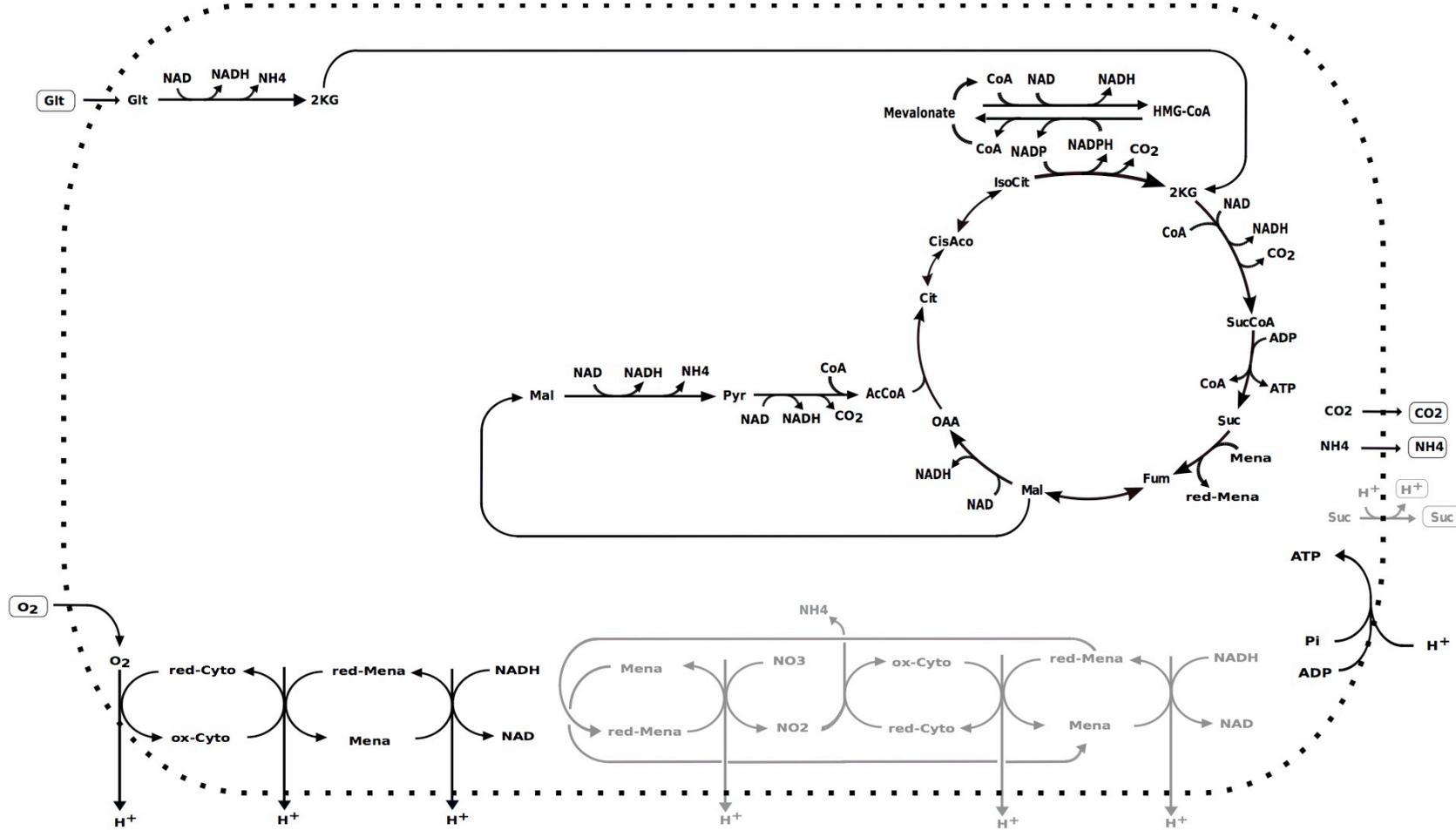


Figure 3-9 ATP production from Glt in the presence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> when the import of Glt was minimised.

When Glt was the only C source available (Figure 3-8), it was metabolised to 2-KG by a NAD-dependent Glt-dh. The 2-KG was further metabolised to Suc by reactions of the TCA cycle, generating ATP and more NADH. The NADH produced was then oxidized in the ETC during aerobic respiration, regenerating NAD. 5.5 molecules of ATP were produced per molecule of Glt consumed in the process. The protons excreted with Suc (but not  $\text{NH}_4^+$ ) contributed to the PMF that pumps the ATP synthase.

When the metabolism was re-directed to optimise Glt consumption for energy production (Figure 3-9), the model initially catabolised Glt in the same manner as when the total net flux through the system was reduced. However, this was used to fuel the whole TCA cycle, utilising part of the Mal generated in the process to obtain more Pyr and AcCoA, thus maintaining flux through the cycle. The NADH produced was then oxidized in the ETC during aerobic respiration, which regenerated NAD and produced more ATP. 20.3 ATP molecules were obtained per molecule of Glt. The protons associated with the  $\text{NH}_4^+$  excreted in this solution did not contribute to the PMF that pumps the ATP synthase. This was assessed by multiplying the number of protons excreted by the reactions of the ETC involved in generating the PMF needed to pump the ATP synthase by the flux they carry in this solution. These number was equivalent to that obtained by multiplying the protons imported by the ATP synthase reaction by its net flux. Therefore, it could be concluded that no other excreted protons contributed to the PMF needed for the action of the ATP synthase in this solution.

These two strategies for Glt utilisation might be of importance under conditions in which Glc and subsequently the main glycogenic amino acids (Ala, Ser, Thr and Gly) have been depleted from the media.

### ***3.3.2.3 ATP production from acetate***

In order to investigate the ability of the system to reproduce the strategy for ATP synthesis described for staphylococci during the second metabolic state (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016), the model was analysed for ATP production reproducing the conditions found during stationary growth phase (Glc depletion and Ac accumulation).

#### **Model analysis**

In order to study the ability of the model to utilise Ac for ATP synthesis, the LP described in Section 3.2.1 was solved again considering minimisation of total flux as the objective function, constraining flux through the Glc medium importer to 0 and temporarily allowing Ac uptake by re-defining its exporter as a temporarily reversible reaction.

#### **Results**

The LP-based analysis solution obtained is described in Figure 3-10.



In the solution obtained in the presence of both electron acceptors when Ac was the sole C source available (Figure 3-10), Ac was imported and fed into the TCA cycle prior conversion to AcCoA in an ATP-consuming reaction catalysed by the AcCoA ligase (EC 6.2.1.13). The NADH generated in the TCA cycle entered the ETC during aerobic respiration, allowing further ATP production and maintaining the redox balance. This strategy presented a P/O ratio of 1.75, for both, the total and the oxidative phosphorylation and achieved production of 7 molecules of ATP per molecule of Ac consumed. Furthermore, it is consistent with the second metabolic state described for staphylococci upon total Glc depletion (Chapter 0, Section 1.4.2.2).

### **3.3.3 Biofilm energy metabolism: production of acetoin and butanediol and their utilisation for ATP synthesis**

Having defined possible ATP production strategies that reproduce physiological behaviours described for staphylococcal planktonic cultures under a range of environmental conditions we now turn to looking at ATP production in the biofilm state. As described in Chapter 0, Section 1.4.4, cells growing in biofilms divert C metabolism to acetoin and butanediol production during the exponential growth phase. This has been suggested to help prevent excessive media acidification caused by the excretion of Ac, Lac and Form and to help restore the redox imbalance caused by Pyr utilisation for Form synthesis by regenerating NAD (Zhu *et al.* 2007). These compounds can subsequently be taken up and utilised to obtain energy during the stationary growth phase (Zhu *et al.* 2007). Interestingly, inhibition of these strategies has been shown to prevent biofilm formation (Cassat *et al.* 2006).

#### ***3.3.3.1 Production of acetoin and butanediol***

##### **Model analysis**

In order to check if the system could reproduce organism's behaviour described above, the model was initially tested for its capability to produce acetoin and butanediol from Glc. Once this was confirmed, it was analysed again for ATP production by a LP set up as previously described in Section 3.2.1 and considering anaerobic conditions. However, the optimisation criteria selected was minimisation of flux through the exporters of charged metabolic by-products, assuming this could be a proxy for minimising acidification of the medium.

##### **Results**

The optimal solution obtained involved Glc fermentation to butanol, a uncharged compound. In this solution, the NAD consumed during glycolysis was regenerated via oxidation of NADH by the 2-hydroxyacyl-CoA-dh and the NAD-dependent menaquinone-oxidoreductase, involved in Pyr fermentation to butanoate. Blockage of flux through either the butanol exporter or the short chain acyl-CoA-dh (EC 1.3.8.1) reaction stopped Pyr fermentation to butanol or butanoate and produced a new optimal solution involving production and excretion of EtHO, another neutral by-product.

### 3.3.3.2 *Utilisation of metabolic by-products for ATP production*

As described in section 1.7.5, staphylococcal cells growing in biofilms ferment Pyr, producing a range of by-products which include Ac, Form, Lac (Resch *et al.* 2005; Zhu *et al.* 2007) and later on, acetoin and butanediol (Yao *et al.* 2005; Xiao *et al.* 2007). These metabolites are then available for use to obtain energy during stationary phase prior conversion to AcCoA (Zhu *et al.* 2007). The work presented in this section studies the system's potential to utilise these by-products for energy production under a range of conditions occurring in biofilms growing on abiotic surfaces in joints: since biofilms consist of several cell layers, it is reasonable to assume that cells in different biofilm depths will encounter different biochemical environments, which would influence their biomass composition and the metabolic processes they perform, leading to a certain level of metabolic cross-feeding between them. Thus, cells in the upper biofilm layers would be exposed to higher O<sub>2</sub> and Glc levels than cells in the bottom layers etc. These situations were reproduced during the analysis described below. A more detailed analysis of the metabolic processes exerted by cells in different biofilm areas and their interactions is out of the scope of this project but will be an interesting work to perform in the future.

#### **Model analysis**

The model is able to produce and export the following metabolic by-products: NH<sub>4</sub><sup>+</sup>, CO<sub>2</sub>, Ac, Form, butanoate, Lac, Suc, EtHO, 2-KG, butanol, butanediol and acetoin. In order to understand which of them, if available, would be preferentially used for ATP production, a LP was set up as previously described in Section 3.2.1, utilising the same Glc-based minimal medium and considering minimisation of total flux as the objective function. By-product-transporting reactions were re-defined, allowing not only excretion but also uptake of these compounds, hence mimicking conditions in which cells could already have conducted fermentation of C sources and the resulting by-products are available for uptake and use for energy production. The LP was modified as corresponded and re-solved in order to explore the system's behaviour under a set of possible states encountered in different areas of biofilms growing in the intra-articular space.

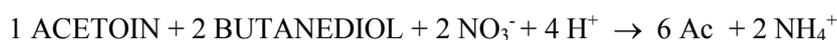
#### **Results**

**a) Cells growing in the top layers of a biofilm:** these cells could still be exposed to O<sub>2</sub>, NO<sub>3</sub><sup>-</sup> and Glc, and potentially, to metabolic by-products produced by cells in deeper layers, hence exerting a fermentative behaviour. Under these conditions, acetoin was the preferred substrate utilised by the system: an optimal solution was obtained in which acetoin was oxidised to AcCoA via action of the acetoin-dh and the acetald-dh, and AcCoA was in turn oxidised to Ac by the Ac-CoA ligase. This process generated ATP and NADH, while reactions of the ETC worked with O<sub>2</sub> as final electron acceptor regenerating NAD and generating further ATP via oxidative phosphorylation. This solution presented the following net stoichiometry:



with 6 ATP molecules produced per molecule of acetoin oxidized.

**b) Cells growing in the middle layers of a biofilm:** these cells would probably be growing under anaerobic conditions and upon Glc depletion, but  $\text{NO}_3^-$  (present in blood and synovial fluid) could still be available if not used-up by cells in the top layers. The solution obtained considering these conditions was similar to the one above, but included and extra step where butanediol was uptaken and oxidised to acetoin via the butanediol-dh. It also involved reactions of anaerobic respiration, with  $\text{NO}_3^-$  being utilised as final electron acceptor. Its net stoichiometry is shown below:



with 19 ATP molecules produced per each molecule of acetoin and pair of butanediol molecules consumed.

**c) Cells growing in the bottom layers of a biofilm:** it is sensible to assume that the environment encountered by cells in this region of the biofilm has been depleted of Glc and, possibly, of both electron acceptors. Upon constraining flux to zero for their importers, the LP solution obtained showed uptake and fermentation of acetoin to butanoate, with AcCoA being produced from acetoin in the same manner as before, now following the same catabolic route as seen for production of ATP from Glc in absence of electron acceptors (Section 3.3.2.1, Figure 3-6). This led to the excretion of butanoate (butyric-acid), with the butyryl-CoA-dh re-oxidising menaquinones in order to allow further oxidation of NADH. These reactions generated a chemiosmotic gradient that fuelled the ATP synthase. The net stoichiometry of this solution is:



with 2 ATP molecules produced per molecule of acetoin consumed.

### 3.3.4 Metabolic responses for production of planktonic biomass under a range of environmental conditions

LP-based analysis was used to investigate the capacity of the model to generate biomass for general quality assessment. The *in silico* medium considered during the analysis was the minimal medium previously defined in section 3.2.1, and the LP was solved several times, mimicking aerobic and anaerobic conditions and presence and absence of  $\text{NO}_3^-$ .

#### Model analysis

The behaviour of the model during biomass production in presence and absence of  $\text{O}_2$  and  $\text{NO}_3^-$  was studied. For this purpose, a LP was defined as described in Equation 2-4 (Chapter 2, Section 2.4.4.1), where the flux through the exporters of biomass components was set to meet the proportions described in the biomass composition of the cell in order to allow production of planktonic biomass. The flux through the ATPase reaction was constrained to meet the growth-associated (GAM) and non-growth

associated (NGAM) maintenance costs. To the best of our knowledge, there is no data currently available regarding ATP requirements for staphylococci. Heinemann *et al.* (2005) considered values of 40 mmol ATP/gDW/h and 5 mmol ATP/gDW/h for the GAM and NGAM costs respectively during LP-based analysis of their *S. aureus* N315 model. Since this compared reasonably well to experimentally determined values for several bacteria (Stephanopoulos 1998), they were also adopted on this study. Specific growth rates for *S. epidermidis* are not available: in the past, authors have assumed values of 1.6 gDW/h for *S. aureus* (Heinemann *et al.* 2005) and 0.91 gDW/h for *S. enterica* sv. Typhimurium (Hartman *et al.* 2014) based on measurements from *E. coli* (Bauchop *et al.* 1960; Feist *et al.* 2007). For simplicity, since NAS are known to typically grow at a slower pace than *S. aureus*, here we assumed a value of 1 gDW/h for *S. epidermidis* RP62A. The results obtained are summarised in Table 3-4. For completeness, a solution for planktonic biomass in the absence of electron acceptors obtained upon blocking flux through the reaction catalysed by the short chain acyl-CoA-dh was also obtained, and the main features of this response are summarised in Table 3-5.

## Results

**Table 3-4 Characterisation of the model behaviour for production of planktonic cell biomass in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> considering minimisation of total net flux through the network as the objective of the analysis**

Electron acceptor available	Reactions carrying flux (n)	Objective value	O <sub>2</sub> uptake	NO <sub>3</sub> <sup>-</sup> uptake	Glc uptake	NH <sub>4</sub> <sup>+</sup> uptake	By-products exported
O <sub>2</sub>	343	397	11.0	0.00	8.50	7.25	Ac, Form, CO <sub>2</sub>
NO <sub>3</sub> <sup>-</sup>	344	398	0.00	6.31	9.29	0.94	Ac, Form, CO <sub>2</sub>
None	346	530	0.00	0.00	19.9	7.25	Form, CO <sub>2</sub> , butanol, butanoate

Unit of the objective value and the uptake rates: mmol/gDW/h. The strategy followed by the system in the presence of both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> is identical to that described in sole presence of O<sub>2</sub>, hence data regarding this solution were omitted.

**Table 3-5 Characterisation of the model behaviour for production of planktonic cell biomass in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> upon blocking flux through the short chain acyl-CoA-dh reaction while minimising total net flux through the network**

Electron acceptor available	Reactions carrying flux (n)	Objective value	O <sub>2</sub> uptake	NO <sub>3</sub> <sup>-</sup> uptake	Glc uptake	NH <sub>4</sub> <sup>+</sup> uptake	By-products exported
None	340	798	0.00	0.00	25.2	7.25	Ac, Form, EtHO, CO <sub>2</sub>

Unit of the objective value and the uptake rates: mmol/gDW/h

## 3.4 Discussion

### 3.4.1 Characterisation of the ETC

Two conclusions can immediately be derived from these results: i) as expected, ATP production with utilisation of  $O_2$  as final electron acceptor (EM1) is a more efficient process than utilisation of  $NO_3^-$ , and yielded a higher ATP/NADH ratio; ii) the reduction of  $NO_2^-$  or  $NO_3^-$  to  $NH_4^+$  (EM2 and EM4) is more efficient than the sole reduction of  $NO_3^-$  to  $NO_2^-$  (EM3). The P/O ratio obtained for EM1 (2.25) falls within the expected value range of 2 to 3 (Garrett 2010; Ferrier 2014) and its associated ATP/NADH ratio (2.25) is in reasonable agreement with the ATP/NADH ratio described for *S. aureus* (Wilkinson 1997; Heinemann *et al.* 2005). The same is true for those modes where anaerobic respiration leads to production of  $NH_4^+$  (EM2 and EM4). Thus these results seem to indicate that flux through the set of reactions included in modes EM1, 2 and 4 would be prioritized over those in EM3 in the biological organism, since these routes produce energy more efficiently and match experimentally determined parameters more accurately. These results prove that the model is capable of conducting aerobic and anaerobic respiration yielding reasonable ATP/NADH and P/O ratios, successfully reproducing this metabolic aspect of the organism.

### 3.4.2 Metabolic responses for ATP under a range of environmental conditions

#### 3.4.2.1 ATP production from glucose

The physiological behaviour observed in staphylococci during exponential growth phase in the presence of  $O_2$  and Glc is consistent with the LP solution that considered minimisation of total flux through the system as optimisation criteria (

Figure 3-3). This confirms that repression of the TCA cycle and diversion of central C flux into Ac production is an optimal strategy for balancing energy production and reduction of the total net flux through the system (proxy for the enzymatic activity associated with the model response (Section 2.2.4)), thus supporting this hypothesis as proposed by others (Molenaar *et al.* 2009; Basan *et al.* 2015). Moreover, this response was reproduced by the system without the need to artificially blocking flux through reactions of the TCA cycle (as was the case during the analyses performed by Heinemann *et al.* in 2005). This demonstrates that metabolic modelling can be used to anticipate the occurrence of mechanisms which implement optimal strategies defined mathematically, such as metabolic control strategies based on the regulation of transcriptional patterns influencing enzymatic catalysis, as occurs during catabolic repression via downregulation of enzymes of the TCA cycle (Blumenthal 1972; Somerville *et al.* 2002; Somerville *et al.* 2003; Deutscher 2008).

Selecting minimisation of Glc import as objective function led to an optimal solution in which the highest ATP yield per molecule of Glc was achieved: involving glycolysis, the TCA cycle and aerobic respiration (

Figure 3-5), this strategy optimises Glc usage for energy production and makes sense, from a metabolic point of view, when the availability of Glc is limited. This accurately reproduces the organism's ability to adapt to conditions where the availability of C sources is limited by obtaining higher energy yields per molecule of C source consumed.

Results also suggested that if both  $O_2$  and  $NO_3^-$  are available,  $O_2$  would be the preferred final electron acceptor of the ETC. This allows for a higher amount of ATP produced per Glc consumed and is in line with previous results described in Section 3.3.1 (elementary mods of the ETC stand-alone module), which showed that ATP production via aerobic respiration was more cost-effective (higher ATP/NADH ratio) than via anaerobic respiration.

In the absence of electron acceptors, the system fermented Glc to Form, Ac and butanoate, excreting Form and butanoate (Figure 3-6). Even though experimental data for *S. epidermidis* reports Glc being fermented mainly to Lac and trace amounts of Ac, Form and  $CO_2$  (Sivakanesan *et al.* 1980), it is possible that the fermentation strategies shown by the organism could be strain-dependent and also vary in response to changes in the environment. The model is capable of fermenting Glc to Lac or to Form, Ac and EtHO, as shown when further constraints were applied, exerting a behaviour more consistent with the *in vitro* observations for staphylococci described to date (Sivakanesan *et al.* 1980; Fuchs *et al.* 2007) (Figure 3-7). These responses seem less efficient, both from an energetic and a protein-salvage point of view but could be consistent with optimal strategies for other unknown biological objectives of the cell (Sivakanesan *et al.* 1980; Fuchs *et al.* 2007).

Some of the responses involving functioning of the ETC showed that the net proton export associated with excretion of charged by-products contributed to the PMF fuelling the ATP synthase (Figure 3-8 and Figure 3-9). This behaviour, although unexpected, could be considered plausible since there is no clear data available on this matter. Further studies on the coupling of the ETC and phosphorylation are required. Moreover, data available regarding the presence and functioning of electrogenic and electroneutral proton/substrate symporters in *S. epidermidis* is scarce and incomplete. Therefore, it is currently not possible to accurately define if or how the transport of these compounds contributes to the PMF and the subsequent phosphorylation process.

### **3.4.2.2 ATP production from glutamate**

Despite not producing a P/O ratio of 1.5 as described in the literature (Tynecka *et al.* 1999; Heinemann *et al.* 2005), the P/O ratios obtained on these analyses were close to this value, especially when the optimisation criteria considered was to minimise Glt consumption. The static nature of the model, which does not take into account kinetic parameters, could explain the differences between

the *in vitro* and the *in silico* results, since LP-based analysis of a structural model is insufficient to consider the variability of the complex biological system. It is also worth noticing that this experimentally obtained P/O ratio of 1.5 was reported without confidence limits, hence its accuracy is difficult to assess. It is however interesting to explore how ATP can be generated from Glt, since this amino acid is, together with Gln, central to N assimilation and the biosynthesis of other amino acids (Reitzer 2003; Somerville 2016) and constitutes the main link between C and N metabolism.

### ***3.4.2.3 ATP production from acetate***

Available experimental data indicate that during the post-exponential growth phase, *S. aureus* converts acetyl-P to AcCoA via the P-acetyl transferase (Pta), although there is insufficient data to confirm the activity of the Ac kinase (ackA) converting Ac to acetyl-P at this stage (Somerville *et al.* 2003), while direct conversion of Ac to AcCoA by the AcCoA ligase was confirmed, as well as AcCoA entering the TCA cycle (Somerville *et al.* 2003). Hence the solution obtained in this analysis, where Ac is metabolised via the AcCoA ligase and the TCA cycle (Figure 3-10), is consistent with the metabolic behaviour described for staphylococci during the post-exponential growth phase.

## **3.4.3 Biofilm energy metabolism: production of acetoin and butanediol and utilisation for ATP synthesis**

### ***3.4.3.1 Production of acetoin and butanediol***

In the optimal solution obtained for ATP production while minimising flux through the exporters of charged by-products Glc was fermented to butanol, an uncharged compound. This response, despite not involving acetoin or butanediol production is optimal for generating energy while minimising the excretion of protons and would indeed help prevent acidification of the media if implemented by the living organism. When extra constraints were applied and flux through either the butanol exporter or the short chain acyl-CoA-dh were blocked, the system fermented Pyr to EtHO, a by-product which excretion again does not involve the excretion of protons. A limitation of the LP-based analysis technique applied here consists in the complexity of solving a LP problem that simultaneously or sequentially considers two or more optimisation criteria (e.g. minimisation of flux through certain exporters and minimisation of total net flux through the system). Furthermore, we currently lack the knowledge to identify the real biological objective followed by an organism and our analysis are based on speculations. Hence based on these results, and despite acetoin and butanediol production not being included in the optimal responses obtained here, their potential involvement in other alternative optimal solutions and in metabolic strategies designed to minimise the export of protons while simultaneously optimising other biological objectives cannot be ruled out.

### 3.4.3.2 Utilisation of metabolic by-products for ATP production

Heterogeneous environments have been described across the structure of bacterial biofilms, with lower layers encountering nutrient restrictions that impair growth, and peripheral layers exhibiting conditions closer to those that fully support planktonic growth (Brauner *et al.* 2016; Dengler Haunreiter *et al.* 2019). The results obtained during the study of ATP synthesis under a range of conditions resembling these chemical environments (Section 3.3.3.2) again exemplified the robustness and adaptability of the organism's metabolic network, providing insight about possible routes for by-product utilisation towards ATP production on RP62A biofilms. This work could be expanded in the future in order to gain further insight into the energy metabolism of biofilms.

### 3.4.4 Metabolic responses for production of planktonic biomass under a range of environmental conditions

The results obtained showed an optimal solution consisting on 343 reactions including transporters (36% of total reactions) required for growth and maintenance in the presence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>. These corresponded to 286 reactions excluding transporters, from which 225 (78.7%) were associated with identifiable genes. A total of 7.25 mmol/gDW/h of N were taken up as NH<sub>4</sub><sup>+</sup> from the medium and were utilised, together with 2-KG, to obtain Glt by the NAD-dependent Glt-dh (EC-1.4.1.2), being later on consumed for the biosynthesis of amino acids. 8.5 mmol/gDW/h of Glc were taken up and used to produce ATP and provide more C for anabolic processes.

As expected, in aerobic conditions the optimal solution obtained did not involve NO<sub>3</sub><sup>-</sup> uptake. This is in line with previous findings where O<sub>2</sub> was identified as the most efficient final electron acceptor for ATP production by calculation of the elementary modes of the ETC (Section 3.3.1). The by-products excreted on this process were Ac, Form and CO<sub>2</sub>. When NO<sub>3</sub><sup>-</sup> was the only electron acceptor available, the solution obtained was similar. Unsurprisingly, when neither O<sub>2</sub> nor NO<sub>3</sub><sup>-</sup> were available, the objective value increased substantially (+33.5%) and so did the Glc demand (+134%).

In all solutions, 2 μmol/gDW/h of niacin were consumed, since this was essential for NAD and NADP production. Its contribution to the final cell N content was so small that could be dismissed. In order to produce biomass, the system unconditionally exported 7 μmol/gDW/h of autoinducer-2 or (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran. Generated in the S-adenosyl-L-methionine cycle, this compound is produced from the pool of L-methionine that is not directly utilised as a protein-building block. This molecule acts as the major methyl donor in the cells (MetaCyc vs 22.6 (Caspi *et al.* 2014)), and is involved in the last step of menaquinone biosynthesis (methylation of demethylmenaquinones). It has also been described as one of the key molecules involved in quorum sensing (Miller *et al.* 2001), being an universal signalling molecule for cell growth both in Gram-positive and Gram-negative bacteria (Zhu *et al.* 2003). Therefore, model analysis provided a metabolic link between cell growth and the excretion of autoinducer-2: since this

compound is generated as a by-product during synthesis of menaquinones and is not metabolised further by the system, it needs to be unconditionally excreted in order to allow production of planktonic biomass *in silico* while complying with the steady state assumption. Therefore, it makes sense that, if autoinducer-2 is excreted as a result of menaquinone synthesis during cell growth, higher levels of this compound in the media will correspond to higher bacterial growth, and, hence higher cell densities in the culture. The concentration of this molecule will eventually reach a certain threshold level, inducing a quorum sensing regulatory response in adjacent cells. This is in line with data describing this molecule as an intercellular signal for cellular density (De Kievit *et al.* 2001; Miller *et al.* 2001; Zhu *et al.* 2003).

### 3.5 Conclusion

The results presented in this chapter demonstrate that the model exhibits strategies which are close to the organism's behaviour *in vitro*. For example, the network is able to produce all biomass components in the experimentally observed proportions described for staphylococci. It is also capable of performing aerobic and anaerobic respiration, yielding reasonable ATP/NADH and P/O ratios. Additionally, investigation of ATP production under a wide set of conditions showed that the system successfully reproduces several other physiological responses, such as the strategies observed in staphylococci during: i) the exponential growth phase in the presence of Glc and O<sub>2</sub>, which involves suppression of the TCA cycle and Ac excretion; ii) the post-exponential growth phase upon Glc depletion, with Ac being imported and fed into the TCA cycle via AcCoA; iii) the fermentative behaviour observed in the absence of electron acceptors; and iv) ATP production from acetoin and butanediol, as has been described on staphylococcal biofilms. Furthermore, the association established *in silico* between biomass production and excretion of autoinducer-2 coincides with the role of this molecule as an intercellular signal for growth. This exemplifies how LP-based analysis of a GSM could provide a mathematical explanation for these type of biological observations, helping us to understand their metabolic basis or implications.

Finally, a very interesting conclusion can be derived from these results: metabolic modelling is able to suggest strategies that organisms would follow in order to optimise certain biological objectives, thus anticipating regulatory and transcriptional patterns (e.g. the oxidation of Glc to Ac in the presence of O<sub>2</sub> and consequent generation of ATP via substrate level phosphorylation as a way to produce energy while minimising total net flux through the system, which the biological organism achieves by repressing the enzymes of the TCA cycle).

In summary, the *in silico* results obtained during LP-based analysis of the system upon a wide set of constraints are highly consistent with the physiological behaviour of the organism, which supports the biological significance of findings derived from analysing this model and validates its use as an *in silico* proxy to study the metabolism of *S. epidermidis* RP62A.

## 4 Model refinement and validation: minimal growth requirements

### 4.1 Introduction

Comparing model-predicted minimal growth requirements with experimental data is a common way to improve and validate GSMs, as exemplified by previous studies in which minimal growth requirements were computed for published GSMs of staphylococci and compared with experimental results for the purpose of model refinement and validation. This chapter describes the acquisition and analysis of experimental data for curation of reactions leading to production of vitamins and amino acids included in the biomass composition of the cell. Moreover, since proteins are key components of certain biofilm types and amino acids seem to be important for biofilm metabolism (Zhu *et al.* 2007), minimising inconsistencies between the model and the organism's capabilities in terms of amino acids biosynthesis and catabolism is key to ensure that *in silico* results can be trusted. In many cases, staphylococcal strains which have been described as auxotrophic under certain circumstances seemed to revert to a prototrophic state upon environmental changes, exemplifying the ability of these organisms to adapt their metabolism to grow on a range of niches and colonise different hosts (Gladstone 1937; Emmett *et al.* 1975; Somerville 2016) using mechanisms not yet fully understood.

Initially, RP62A was tested *in vitro* for its ability to produce biomass upon removal of single amino acids and vitamins from a chemically defined medium. The impact of nutrient deprivation on biofilm formation was also assessed. The laboratory strain was checked for possible mutations which could have had affected these *in vitro* results. The genome of the modelled organism was thoroughly examined in order to define the biosynthetic potential of the strain and the model was modified accordingly, thus minimising the impact of excessive gap-filling propagated from construction of the PGDB or completing the network with reactions initially absent. The model was then re-analysed for biomass production. Finally, the effect of removing single compounds from the *in silico* medium on the objective value and the Glc consumption was calculated with respect to the solution obtained on the un-modified medium. These datasets were considered together and used to perform several rounds of manual curation. After this, the system was re-analysed and the results obtained were used to design new validation experiments presented in Chapter 5.

#### 4.1.1 Use of experimental data to validate GSMs of staphylococci

Experimentally-defined auxotrophies and minimal growth requirements identified with GSMs of *S. aureus* N315 are summarised and compared in Table 4-1 (below) and described in detail in this section:

**Table 4-1 Comparison of minimal growth requirements for *S. aureus* N315 identified by analysis of four different GSMs**

GSM	Strain	Auxotrophies identified by analysis of the GSM	Essential amino acids documented experimentally and reported by these authors	Biosynthetic pathways lacking in the genome reported by these authors
iMH551 Heinemann <i>et al.</i> (2005)	<i>S. aureus</i> N315	None (Heinemann <i>et al.</i> 2005)	None (Gladstone 1937)	None (Kuroda <i>et al.</i> 2001)
iSB619 Becker and Palsson (2005)	<i>S. aureus</i> N315	Growth without an additional N source is possible if any one of the following amino acids or derivatives is provided: Ala, Arg, Asp, Glt, Gly, Pro, Ser, Thr or ornithine (Becker <i>et al.</i> 2005).	Ala, Arg, Gly, Ile, Pro and Val (Kuroda <i>et al.</i> 2001)	None (Kuroda <i>et al.</i> 2001)
Lee <i>et al.</i> (2009)	<i>S. aureus</i> N315	Pro and Ser (Lee <i>et al.</i> 2009)	Arg, Cys, Glt, Leu, Phe, Pro, Thr, Val (Rudin <i>et al.</i> 1974)	Pro and Ser (although a bypass is present to obtain Pro from ornithine (Lee <i>et al.</i> 2009)).
Bosi <i>et al.</i> (2015)	<i>S. aureus</i> N315	Leu (Bosi <i>et al.</i> 2016)	Arg, Asn, Cys, His, Leu, Met, Phe, Pro, Ser, Thr, Trp and Tyr (Bosi <i>et al.</i> 2016)	None (Kuroda <i>et al.</i> 2001)

Auxotrophies reported by Lee *et al.* (2009) are general for drug-resistant *S. aureus* and not specific for *S. aureus* N315.

In 2005, Heinemann *et al.* generated a GSM for the *S. aureus* strain N315, which genome annotation contains completed biosynthetic pathways for all amino acids (Kuroda *et al.* 2001; Heinemann *et al.* 2005). Accordingly, this GSM was capable of simulating bacterial growth without their supplementation. The model predicted an auxotrophy for niacin (needed for the synthesis of NAD and NADP), while biotin and thiamine, which had been described as essential vitamins for *S. aureus* in the literature, represent prosthetic groups and were not directly involved in any reaction leading to biomass production. Another GSM for *S. aureus* N315, constructed by Becker and Palsson in 2005, described the following compounds as feasible single N sources for growth: Ala, Arg, Asp, Glt, Gly, Pro, Ser, Thr and ornithine (Becker *et al.* 2005). From those, four (Ala, Arg, Gly and Pro) have been reported as essential *in vitro* (Kuroda *et al.* 2001). The vitamins thiamine and niacin were defined as essential by both the model and experimental results, while biotin only appeared to be essential *in vitro*.

Four years later, Lee *et al.* generated a new GSM for N315 and adapted it for 12 other *S. aureus* strains (Lee *et al.* 2009). After comparing *in vitro* minimal growth requirements for these strains with a previously published minimal medium for staphylococci (Rudin *et al.* 1974), the study concluded that Pro could not be synthesised by any of them and had to be supplemented in the media together with Ser, which synthetic pathway appear to be incomplete. Regarding essentiality of vitamins, a synthetic pathway was found for biotin (from pimelate) but not for niacin or thiamine, which were

then included in the *in silico* minimal medium.

Finally, in 2015, Bosi *et al* constructed GSMs for 64 *S. aureus* strains (including N315) and used them to study a possible effect of the genomic presence and absence of amino acid's biosynthetic pathways in niche adaptation (Bosi *et al.* 2016). Analysis of these models concluded that all strains needed thiamine and niacin to grow in Glc-based minimal media. Analysis of the GSM for strain N315 indicated that in the presence of NH<sub>3</sub>, both NH<sub>3</sub> and the amino acid Leu were uptaken for biomass production. Several other strains presented strain-specific needs for amino acids, vitamins and even nucleotides. These authors reported growth *in vitro* for N315 in a M9-based minimal medium containing Arg, Asn, Cys, His, Met, Phe, Trp and Tyr and further supplemented with Leu, Pro, Ser, Thr and the vitamins thiamine and niacin.

The discrepancies observed when minimal growth requirements were defined for the same organism by analysing different GSMs (Table 4-1) could be explained as a consequence of the utilization of different metabolic databases and modelling techniques for model construction and analysis, which in turn involve different degrees of automatic-gap filling, and the thoroughness of the subsequent manual curation process applied to these models. Generally speaking, a certain degree of variation between model predictions and experimental results is expected, and acceptable as long as these discrepancies are thoroughly investigated and can be explained in a reasonable manner (e.g. as a result of gene regulation).

#### **4.1.2 Comparison between minimal growth requirements for *S. epidermidis* RP62A defined *in vitro* and by LP-based analysis of the GSM**

LP-based analysis of the model for biomass production after fundamental curation defined auxotrophies for Asn, Cys, Met and Phe (Chapter 2, Section 2.4.4.5). Due to the high degree of variability when reporting auxotrophies for staphylococci (Knight 1937; Emmett *et al.* 1975) it was difficult to discern if these represented real biological auxotrophies or reflected errors or gaps in the network based solely in comparison with the unique experimental dataset available for RP62A (Hussain *et al.* 1991) described in Chapter 0, Section 1.5.1. Therefore, new experimental data on this matter was obtained and used for further model refinement together with exhaustive examination of the genome content and identification of possible mutations in the laboratory strain used in this project.

The table below (Table 4-2) summarises the level of agreement between model-predicted auxotrophies for amino acids, published experimental data and the *in vitro* findings obtained during the work presented on this and the following chapter (Chapter 5). Note that after curation, auxotrophies described *in silico* completely match those defined *in vitro*:

**Table 4-2 Comparison between the biosynthetic potential for amino acids encoded in the genome of *S. epidermidis* RP62A, experimentally reported auxotrophies and auxotrophies defined by model analysis.**

Amino acid	Absence of biosynthetic genes in RP62A	Auxotrophy reported by Husain <i>et al.</i> (1991)	Auxotrophy experimentally determined in this study	Essentiality according to model analysis before curation
Ala	No	No	No, but its absence delays growth.	No
Arg	No, however, the urea cycle is broken, which could affect regeneration of ornithine from Arg.	Yes	No, but its absence delays growth.	No
Asn	No	No	No	Yes
Asp	No	No	No	No
Cys	No	Yes	No	Yes
Glt	No	No	No, but its absence delays growth.	No
Gly	No	No	No	No
His	No	No	No	No
Ile	No	No, but its absence delays growth.	No	No
Leu	No	No, but its absence delays growth.	No, but its absence delays growth.	No
Lys	No	No	No	No
Met	No	No	No	Yes
Phe	No	No	No	Yes
Pro	No: <i>de novo</i> synthesis from Glt is absent, however, a biosynthetic bypass through ornithine is present.	No	No	No
Ser	No	No	No	No
Thr	No	No, but its absence delays growth.	No, but its absence delays growth.	No
Tyr	No	No, but its absence delays growth.	No	No
Trp	No	Yes	No, but its absence delays growth.	No
Val	No	Yes	No, but its absence delays growth.	No

After curation, none of the amino acids were essential according to model analysis results.

Discrepancies between the two experimental datasets are believed a result of pathway repression caused by regulatory events and so is the *in vitro* absence of growth without Pro observed in this study.

## 4.2 Methods

### 4.2.1 Experimental assessment of minimal growth requirements in *S. epidermidis* RP62A

#### 4.2.1.1 Minimal media composition

A minimal medium for RP62A, referred here as MM medium, was defined in the following manner: the concentration of salts, trace elements, amino acids and vitamins (thiamine, niacin and biotin) were based on those utilised by Hussain *et al.* for the preparation of HHW medium (a chemically-defined rich medium optimized for growth and biofilm formation of NAS containing 18 amino acids) (Hussain *et al.* 1991). Gln was not added to the MM medium recipe, since it can typically be obtained from Glt and was not present in HHW medium. However, Asn, which was also absent in the HHW medium recipe, was included in the MM medium formulation, since analysis of the GSM initially predicted an auxotrophy for it (data not shown). The concentration of Glc utilised was reduced to 11.11 mmol (0.2%) in comparison to than in HHW medium (55.55 mmol (1%)), since higher Glc concentrations (1%, 2.7%, 5% or 10%) have been reported to encourage phenotypic changes and induce biofilm formation (Lim *et al.* 2004; Otto 2008; Agarwal *et al.* 2013; Fernanda Cristina Possamai Rossatto 2017). Other compounds present in HHW medium but not defined as essential by analysis of the GSM or the literature (i.e. adenine sulphate, guanine hydrochloride and other vitamins) were not included on the MM medium formulation. Note that throughout this document, the term ‘MM medium base’ refers to the composition of the standard MM medium described below (Table 4-3) without amino acids. In order to minimise variation between media batches, large stock solutions of medium components were prepared at the following concentrations in comparison to their final concentration in MM medium: i) salt components (x 100), ii) Glc (x 20), iii) amino acid mix in the standard medium (x 25), iv) individual amino acid stocks (x 25), v) vitamins (x 1000). These were stored according to the manufactures instructions and brought back to room temperature (when corresponded) before preparing fresh media at the beginning of each experiment.

**Table 4-3 Composition of the standard MM medium formulated for minimal growth requirement experiments**

<b>MM medium composition (per L)</b>	<b>mmol</b>
Ammonium iron(II) sulphate hexahydrate ( $(\text{NH}_4^+)_2\text{SO}_4 \cdot \text{FeSO}_4 \cdot 6\text{H}_2\text{O}$ )	0.21
Calcium chloride anhydrous ( $\text{CaCl}_2$ )	0.45
Disodium hydrogen phosphate anhydrous ( $\text{Na}_2\text{HPO}_4$ )	56.1
Magnesium sulphate anhydrous ( $\text{MgSO}_4$ )	2.03
Manganese sulfate monohydrate ( $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ )	0.29
Potassium dihydrogen phosphate ( $\text{KH}_2\text{PO}_4$ )	22.0
Glc	11.1
L-Ala	1.12
L-Arg	0.57
L-Asn	1.13
L-Asp	1.13
L-Cys	0.41
Gly	1.33
L-Glt	1.02
L-His	0.64
L-Ile	1.14
L-Leu	1.14
L-Lys	0.68
L-Met	0.67
L-Phe	0.60
L-Pro	1.30
L-Ser	0.95
L-Thr	1.26
L-Trp	0.49
L-Tyr	0.55
L-Val	1.28
Biotin	0.04
Nicotinic acid (niacin)	0.16
Thiamine hydrochloride	0.59
Final pH $7.2 \pm 0.2$ at $25^\circ\text{C}$ (adjusted with NaOH 0.1M)	

### **4.2.1.2 Inoculum and bacterial strains**

The *S. epidermidis* RP62A strain used for the experimental work of the entire project was purchased from NCTC. A 25% glycerol stock of this strain was streaked out on BHI agar (an undefined rich medium) and incubated overnight at 37C. 10 ml of fresh BHI broth were inoculated with 3 individual colonies and incubated for 18 hours (overnight) at 37C, shaking (180 rpm). Cells from the overnight culture were recovered by centrifugation at 3000 rcf for 5 minutes. Bacterial pellets were washed twice with sterile PBS in order to remove carry over of essential nutrients and finally re-suspended with PBS up to the original sample volume. An approximate concentration of  $1.5 \times 10^8$  CFU/ml for the bacterial inoculum was estimated by serial dilution and cell counting.

### **4.2.1.3 Experimental setup**

Batches of media for growth tests were prepared in the following manner: MM medium from which each individual vitamin or amino acid was omitted at a time, hence refereed here as MM<sup>-</sup> medium, was used to test individual auxotrophies. BHI broth was used to monitor growth in a rich medium. Unmodified MM medium was used as a positive control for growth in the standard formulation. Non-inoculated MM broth and BHI broth were used as sterility controls. Finally, two more test samples were included in order to obtain additional data for model refinement: MM<sup>-</sup> medium lacking a combination of Cys and Met, for the study of sulphur metabolism, and MM<sup>-</sup> medium lacking Asn and Asp, for rigorous investigation of Asn biosynthesis, which was initially described as essential by LP-based analysis of the model.

A cell inoculum was obtained as described above and added in a 1/100 proportion to 2 ml of each test medium, given an approximate final cell concentration of  $1.5 \times 10^6$  CFU/ml in the test samples. 150 ul of each sample were then added to three independent wells in a 96 well plate and incubated at 37°C and 180 rpm for 48 hours. Cell growth was monitored by measuring optical density spectrophotometrically at 600 nm ( $A_{600}$ ) at times 0, 4, 6, 8, 24, 25, 26 and 48 hours. Cultures were grown in triplicates and three individual 96 well plates were set up as explained above. The resulting growth data were expressed as the means from these three independent biological replicates, with a total of 9 individual measurements per growth condition. Background levels were calculated by averaging values obtained for the sterility controls in BHI and MM medium and were then subtracted from the absorbance measurements of the tests samples as corresponded. For the sole condition where removal of a single amino acid (Ala) led to an apparent growth delay combined with and a lower maximum  $A_{600}$  value than that measured for cultures in the standard MM medium, the statistical significance of this reduction in maximum growth was checked by performing a Student's t-test.

## 4.2.2 Assessment of the impact of amino acid deprivation on biofilm formation

At the end of the experiment described in Section 4.2.1, the level of biofilm formation on the cultured plates was assessed with the crystal violet staining method in order to investigate the effect of amino acid deprivation on biofilm formation:

### 4.2.2.1 Crystal violet biofilm staining method

A modified microtiter plate-based method for the staining and quantification of biofilm biomass was used. This method is based upon incubation of NAS cultures in 96 well plates as previously described by other authors (Christensen *et al.* 1985; Stepanovic *et al.* 2007; Baldan *et al.* 2012) and the protocol followed was carefully optimised before the start of this project in order to enable reproducible growth and biofilm formation in NAS. In summary, at the end of the incubation period (Section 4.2.1.3) the wells in the cultured plates were gently washed twice with PBS to ensure removal of planktonic cells. The remaining bacterial biomass that continued to be adhered to the well surfaces was considered biofilm biomass and was fixed with 100% ethanol for 15 minutes. Then, the ethanol was discarded and the biofilms were air-dried before being stained with a 2% crystal violet solution for 5 minutes. After this, excess dye was removed and the dye that remained staining the biofilm biomass was re-solubilised in 33% glacial acetic acid before measuring its absorbance at 595nm. The optical density readings obtained were directly proportional to the amount of dye retained by the biofilm biomass present on the wells, and therefore, directly proportional to the amount of biofilm produced. Thus the  $A_{595}$  values obtained upon staining of the biofilms were used to calculate the relative level of biofilm formation in each sample. For this, the absorbance values were normalised to 100 with respect to the highest value measured on the plates. The effect of single amino acid removal on biofilm formation was visualised by plotting these values on a bar diagram and the variations observed with respect to the biofilm formation level exhibited by cultures growing in the standard MM medium were calculated and expressed as percentages (Appendix C, Section 9.3.2, Table 9-3).

## 4.2.3 Comparison between experimental data and *in silico* results

### 4.2.3.1 Model analysis for essentiality of media components

The LP-based analysis technique was used as described in Chapter 2, Section 2.4.4.1 and Appendix A, Section 9.1.2.7 to check for essentiality of media components. Initially, the model failed to produce several of the biomass components (described in Section 2.4.3.4), exhibiting auxotrophies for niacin and the amino acids Asn, Cys, Met and Phe. These results were compared with the experimental data and the discrepancies identified were thoroughly investigated. Finally, the model was modified as corresponded and re-analysed. This process was repeated until the system's

behaviour agreed with the conclusions of the investigations performed, thus refining the GSM, and the remaining discrepancies could be explained in a reasonable manner. Extensive manual curation finally led to the model producing each individual biomass component from an *in silico* minimal medium composed solely of Glc, core set substrates, H<sub>2</sub>O and supplemented with the vitamin niacin.

#### ***4.2.3.2 Study of the biosynthetic potential encoded in the genome of the organism***

The presence or absence of key metabolic enzymes encoded in the genome of the reference strain from which the model is derived was studied as follows: when identified, the absence of reactions causing discontinuities in the network was corroborated by interrogation of the databases KEGG and BRENDA and the relevant literature. If, according to any of these sources, relevant enzymes catabolising missing reactions were reported as present on *S. epidermidis* RP62A, their corresponding amino acid sequences were retrieved from BRENDA and the Artemis software (release 16.0.0) (Carver *et al.* 2011) was then used to search for possible matching sequences in the genome of the reference strain. If matches were found, thus evidencing the existence of the enzymes in the organism, the reactions catalysed by them were included in the model.

#### ***4.2.3.3 Identification of possible mutations in the laboratory strain***

A possible source for discrepancies between the behaviour of the model and the experimental results could be the incidence of mutations on genes encoding key biosynthetic enzymes in the laboratory strain with respect to the reference strain. In order to explore this possibility, the presence of genomic differences between the RP62A reference strain and the laboratory strain was investigated. The genome sequence of the reference strain was downloaded from RefSeq (NCBI). Whole-genome sequencing data for the laboratory strain was obtained with Illumina<sup>®</sup> NextSeq by other members of the team (Claire Hill and Emma Manners). Both genome sequences were automatically annotated with Prokka (v. 1.11) (Seemann 2014) and subsequently compared with Roary: the pangenome pipeline (Page *et al.* 2015), applying default parameters. Amino acid sequences of biosynthesising enzymes of interest were extracted from either BRENDA (Jeske *et al.* 2018), BioCyc (Karp *et al.* 2017) or KEGG (Kanehisa *et al.* 2000). ACT: the Artemis Comparison Tool (Carver *et al.* 2005) was then used to identify and extract the corresponding amino acid sequences in both the reference and the laboratory strain and the incidence of possible mutations in the laboratory strain with respect to the reference strain was assessed with SeaView: a multiplatform graphical user interface for sequence alignment (v. 4) (Gouy *et al.* 2009). Finally, the Pfam protein families database (v. 32.0) (El-Gebali *et al.* 2018) was used to identify important areas of protein activity, indicating sequence regions where the incidence of mutations could alter or prevent the functionality of the enzyme. No significant mutations were found on any of the enzymes investigated.

#### **4.2.3.4 Model refinement**

Model requirements for biomass production defined by LP-based analysis were compared with the experimental data: for those amino acids defined as essential *in silico* and non-essential *in vitro* exhaustive investigation of the RP62A genome was conducted, focusing on their corresponding biosynthetic pathways. When all genes were present in the genome but reactions were missing in the network these were added to the model, which was then re-analysed for biomass production. If the compound was still not produced by the system, the stoichiometry and directionality of reactions potentially leading to its synthesis were checked and corrected as corresponded, following the methods described in Section 2.4.4.4. However, if an auxotrophy was detected experimentally for a certain compound but its full biosynthetic pathways was present in the genome and the model, the sequences of these biosynthetic enzymes were analysed for the incidence of mutations in the laboratory strain that could explain the *in vitro* results (Section 4.2.3.3). As a result of this work, several reactions were introduced, modified or removed from the system. Specific details can be found in Appendix D, Section 9.4.

#### **4.2.3.5 Computation of the effect of single amino acid deprivation on biomass production by LP-based analysis**

LP was used to analyse the effect of single amino acid deprivation on biomass production when compared to biomass production in a standard *in silico* MM medium: the model was initially analysed for production of cell biomass while satisfying the GAM and NGAM cell demand in a similar fashion as described in Chapter 3, Section 3.3.4. The LP was re-solved blocking the import of either  $\text{NH}_4^+$  or each single amino acid at a time, as well as in the absence of all amino acids. The effects caused on: i) the objective value (total net flux through the system (mmol/gDW/h)) and ii) Glc consumption (mmol/gDW/h) were calculated for all solutions. For simplicity, the results were filtered so only changes in reaction fluxes  $> 1e^{-3}$  mmol/gDW/h were taken into consideration. The *in silico* MM medium includes the same components as the medium used *in vitro* with the exception of salts and the vitamins biotin and thiamine since they are not involved in any biosynthetic reactions in the system or considered on the biomass composition of the cell. The model file used for this work was Sepi\_MM.spy.

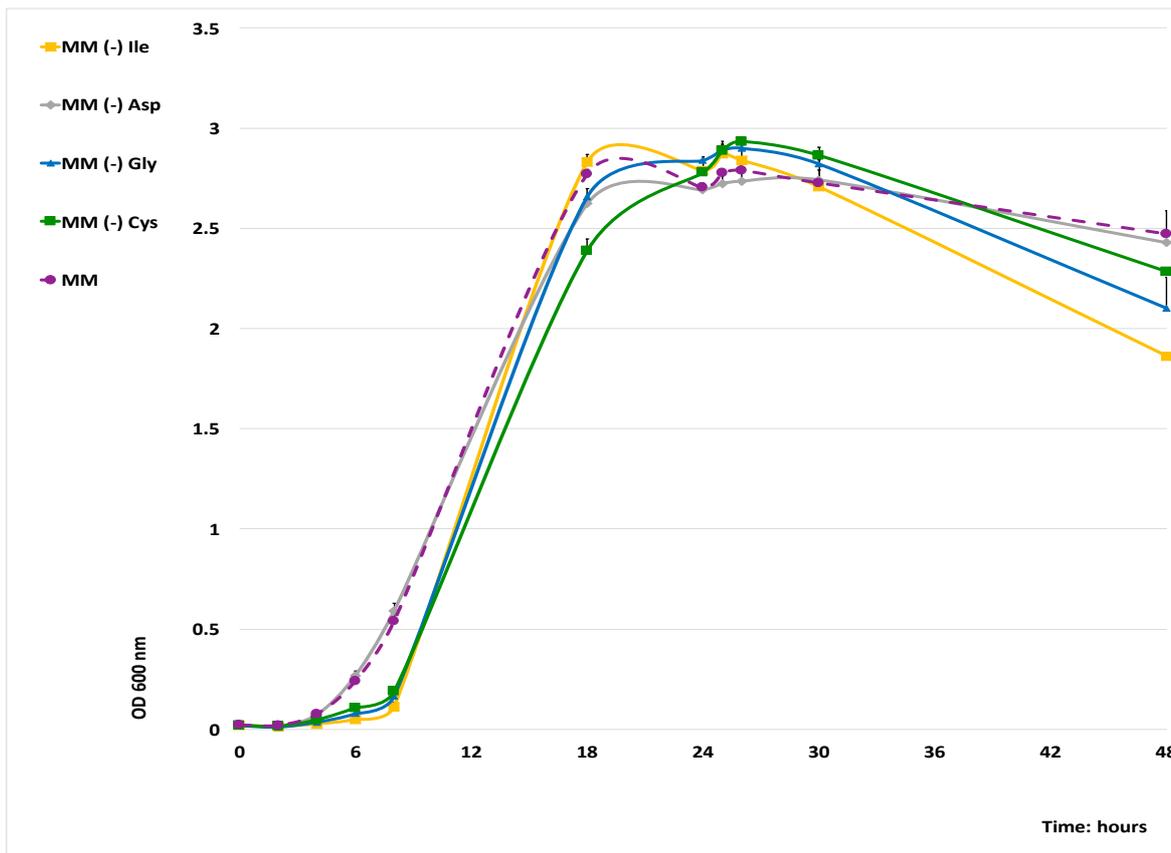
Relevant code to the work described in this chapter can be found in Appendix A, Sections 9.1.2.4, 9.1.2.5, 9.1.3.1 and 9.1.3.2.

## 4.3 Results

### 4.3.1 Experimental assessment of requirements for vitamins and amino acids in *S. epidermidis* RP62A

Although experimental growth requirements for RP62A had been previously described by Hussain *et al.* in 1991 (Chapter 0, Section 1.5.1), the lack of whole-genome sequencing data available for that specific strain makes it impossible to determine if it presented any genetic differences with the RP62A reference strain from which this GSM is derived. As this could be a potential cause of variation on the organism's growth phenotype, minimal growth requirement experiments were repeated with a *S. epidermidis* RP62A strain purchased from NCTC (the laboratory strain for this project). For completion, and in order to identify any possible genetic differences between this and the reference strain, their genome sequences were compared (Section 4.3.2).

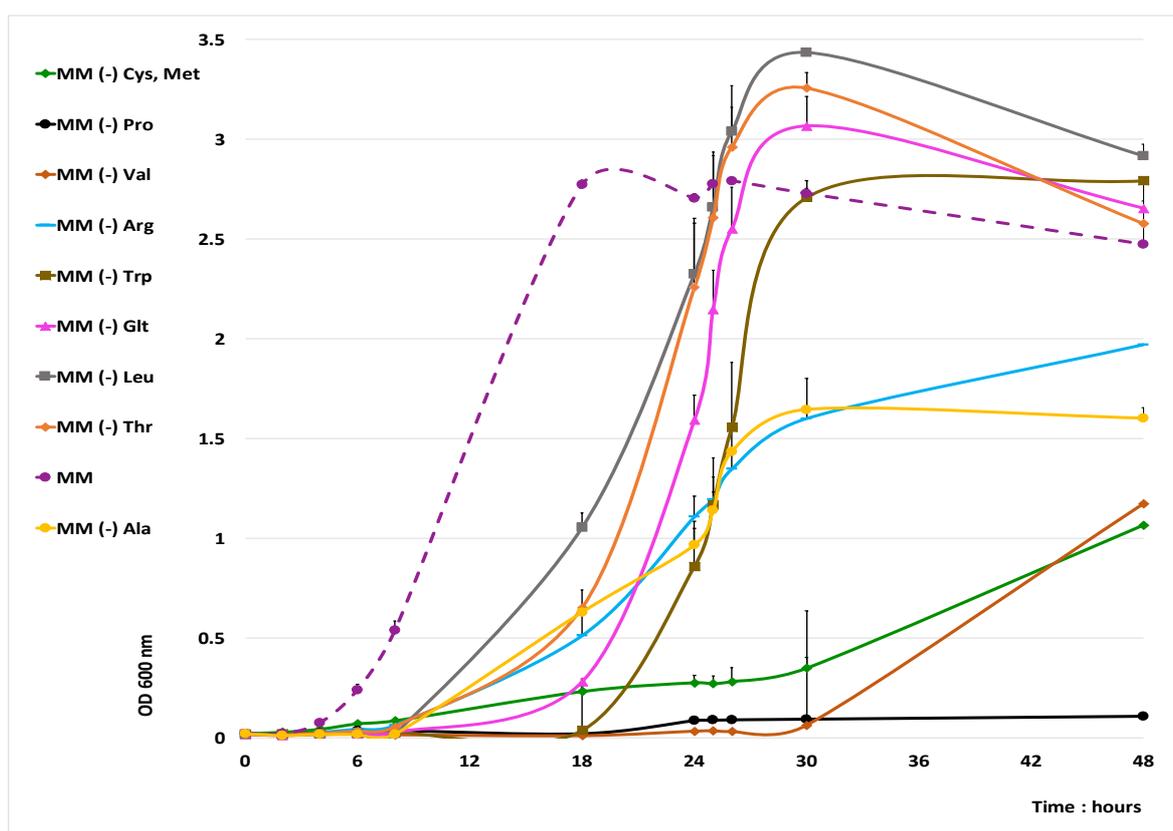
The *in vitro* minimal growth requirements for RP62A were investigated as described in Section 4.2.1. The effect caused by deprivation of single amino acids on bacterial growth can be summarised as follows: inspection and measurement of  $A_{600}$  values of the samples over time showed clear differences in growth between the media tested (Figure 4-1 and Figure 4-2) and are presented in Table 4-4. The effect on growth caused by deprivation of single vitamins is shown in Appendix C, Section, 9.3.1, Figure 9-2.



**Figure 4-1 Growth curves for *S. epidermidis* RP62A in four representative modified MM<sup>-</sup> medium samples presenting no apparent delay or a slight delay in growth in comparison to cultures on standard MM medium.**

Legend: purple dashed line (MM) = standard MM medium; yellow line (MM (-) Ile) = MM medium without Ile; grey line (MM (-) Asp) = MM medium without Asp; blue line (MM (-) Gly) = MM medium without Gly; green line (MM (-) Cys) = MM medium without Cys. Each data point corresponds to the mean  $A_{600}$  value of three independent biological replicates. Error bars = SEM.

For practical reasons, Figure 4-1 above shows growth data for four representative test media samples presenting no apparent delay in growth or a slight growth delay when compared with growth in the standard MM medium. A similar figure can be found in Appendix C, Section 9.3.1, Figure 9-1, including growth curves for BHI and all modified MM<sup>-</sup> medium samples presenting a similar pattern to those in Figure 4-1. These corresponded to cultures grown in MM<sup>-</sup> medium lacking the following individual amino acids: Asn, Asp, Cys, Gly, His, Ile, Lys, Met, Phe, Ser, Tyr and both Asn and Asp.



**Figure 4-2** Growth curves for *S. epidermidis* RP62A in all modified MM medium samples exhibiting an apparent growth delay in comparison to cultures on standard MM medium.

Legend: purple dashed line (MM) = standard MM medium; green line (MM (-) Cys and Met) = MM medium without Cys and Met; black line (MM (-) Pro) = MM medium without Pro; light brown line (MM (-) Val) = MM medium without Val; blue line (MM (-) Arg) = MM medium without Arg; dark brown line (MM (-) Trp) = MM medium without Trp; pink line (MM (-) Glt) = MM medium without Glt; grey line (MM (-) Leu) = MM medium without Leu; orange line (MM (-) Thr) = MM medium without Thr; yellow line (MM (-) Ala) = MM medium without Ala. Each data point corresponds to the mean  $A_{600}$  value of three independent biological replicates. Error bars = SEM.

**Table 4-4** Summary of the *in vitro* growth effects observed upon removal of single amino acids from the MM medium compared with the experimental observations produced by Hussain *et al.* (1991).

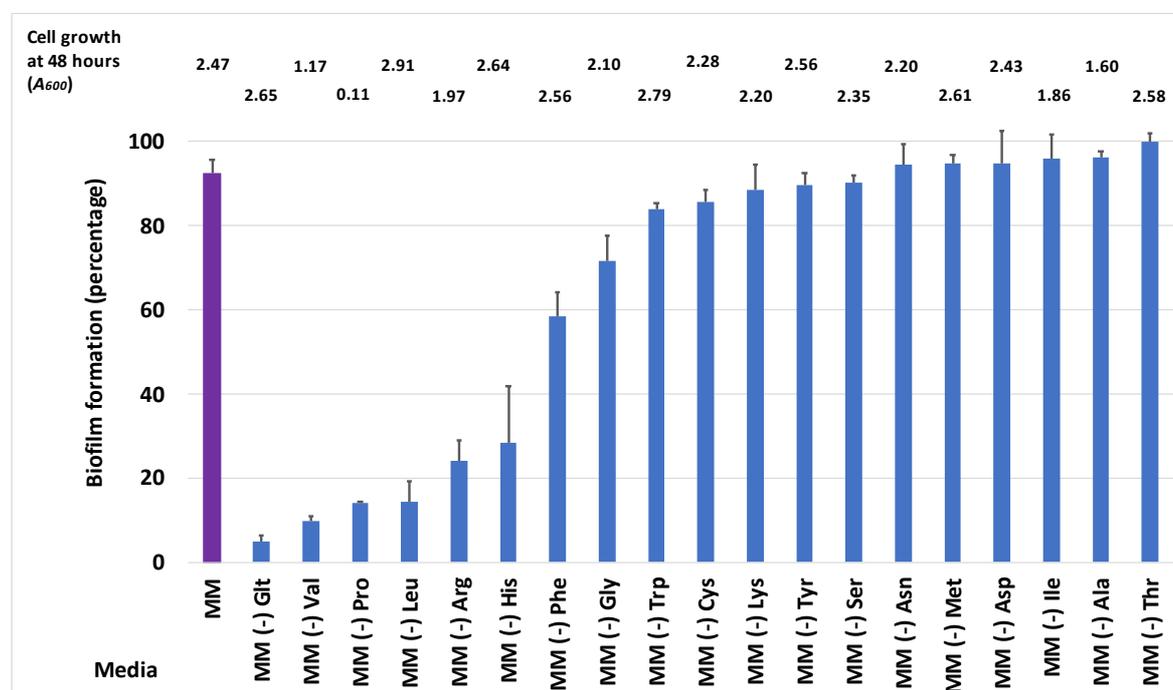
Effect of single amino acid deprivation	Growth apparently delayed or absent by 18 hours	No apparent growth by 24 hours	No apparent growth by 48 hours
Hussain <i>et al.</i> (1991)	Arg, Cys, Ile, Leu, Thr, Trp, Tyr, Val	Non-reported	Arg, Cys, Trp, Val
This study	Ala, Arg, Glt, Leu, Pro, Thr, Trp, Val	Pro, Val	Pro

There are multiple valuable parameters of bacterial growth curves (Zwietering *et al.* 1990) to be considered when studying bacterial growth. Some of these are: i) the maximum  $A_{600}$  value of the growth curve or maximum measurement of growth for a given culture; ii) the specific growth rate, given by the slope of the line fitting the growth curve section that is approximately linear during the exponential growth phase; and iii) the duration of the lag phase, or the length of the incubation period

before growth begins, in which cells are active but not yet dividing. In this experiment, no growth was detected in the sterility controls for BHI and MM medium. Cells in standard MM medium presented a maximum  $A_{600}$  value of  $2.79 \pm 0.05$  SD;  $CI_{95\%} = [2.89, 2.68]$ , a specific growth rate of  $0.23 \text{ h}^{-1} \pm 0.02 \text{ h}^{-1}$  SD;  $CI_{95\%} = [0.26, 0.19]$ ) and a lag phase lasting between 4 and 6 hours. From those cultures exhibiting an apparent delay in growth in comparison to standard MM medium, MM medium without Ala also plateaued at a  $A_{600}$  value ( $1.65 \pm 0.28$  SD;  $CI_{95\%} = [2.20, 1.10]$ ) significantly lower at the 95% confidence level ( $P=0.0023$ ) than those in standard MM medium ( $2.79 \pm 0.05$  SD;  $CI_{95\%} = [2.89, 2.68]$ ). By time 48 hours, only samples in MM medium lacking Pro still presented a lack of growth, with an  $A_{600}$  of  $0.11 \pm 0.00$  SD.

### 4.3.2 Impact of amino acid deprivation on biofilm formation

Removal of single amino acids from the MM medium caused changes not only in the kinetics of cellular growth but also in the cellular phenotype. Specifically, the effect of nutrient deprivation on biofilm formation was assessed at the end of the incubation period of samples being monitored during the study of growth requirements for RP62A (Section 4.3.1): following the method described in Section 4.2.2, the relative levels of biofilm formation were obtained for cultures in standard MM medium and in modified MM medium (Figure 4-3) and are summarised in Appendix C, Section 9.3.2, Table 9-3.



**Figure 4-3** Relative levels of biofilm formation in *S. epidermidis* RP62A cultures growing on standard MM medium and modified MM medium lacking single amino acids

Legend: values correspond to the mean of three independent biological replicates per test condition and have been normalised in a scale 0-100, taking the highest  $A_{595}$  value measured on these experiments as 100. Legend: purple bar (MM) = biofilm formation in standard MM medium; blue bars (MM (-) 'compound name') = biofilm formation in MM medium without 'compound name'. Error bars = SEM. For reference, the  $A_{600}$  values corresponding to these cultures prior staining

of the biofilm biomass are displayed at the top of the figure.

The effect of amino acid deprivation on cell growth and biofilm formation are compared below (Table 4-5):

**Table 4-5 Summary of the effects caused by removal of single amino acids from the standard MM medium on *S. epidermidis* RP62A cell growth and cell phenotype.**

Effects caused by removal of single amino acids		Effect on cell growth	
		Apparent delay	No apparent delay
Variation in biofilm formation with respect to the standard MM medium	Increase	Ala, Thr	Asn, Asp, Ile, Met
	Decrease	Arg, Glt, Leu, Pro, Trp, Val	Cys, His, Lys, Phe, Gly, Ser, Tyr

### 4.3.3 Comparison between experimental data and *in silico* results

#### 4.3.3.1 Model validation and refinement

As a result of the work described in Section 4.2.3, several reactions were introduced, modified or removed from the system. The data obtained from these investigations is compared and summarised in Table 4-6 for amino acids and in Appendix C, Section 9.3.1, Table 9-2 for vitamins. A more detailed description of the outcome of this study, the consequent changes applied to the model and its behaviour before and after curation is provided in Appendix D, Section 9.4.1.

**Table 4-6 Comparison between the biosynthetic potential for amino acids encoded in the genome of *S. epidermidis* RP62A, amino acids auxotrophies reported experimentally and proposed explanations for the discrepancies encountered.**

Amino acid	Absence of biosynthetic genes in RP62A	Auxotrophy reported by Hussain <i>et al.</i> (1991)	Auxotrophy experimentally determined in this study	Proposed explanation for discrepancies between experimental datasets
Arg	No, however, the urea cycle is broken, which could affect regeneration of ornithine from Arg.	Yes	No, but its absence delays growth.	Presence of a biosynthetic pathway for Arg and absence of obvious mutations on these genes indicate that the lack of growth without Arg detected by Hussain <i>et al.</i> is likely to be an effect of regulation.
Cys	No	Yes	No	Normal growth observed without Cys and the presence of a biosynthetic pathway for it indicates lack of essentiality.
Pro	No: <i>de novo</i> synthesis is absent, however, a biosynthetic bypass through ornithine is present.	No	Yes	Presence of a biosynthetic bypass and detection of growth without Pro by Hussain <i>et al.</i> (1991) and later experiments (Chapter 5, Section 5.3.2) indicate that the lack of growth detected on this study is likely due to regulatory processes or the impossibility of the cells to use the ornithine bypass under these conditions.
Trp	No	Yes	No, but its absence delays growth.	Presence of a biosynthetic pathway for Trp and absence of obvious mutations on these genes indicate that the lack of growth without Trp detected by Hussain <i>et al.</i> (1991) is likely to be an effect of regulation.
Val	No	Yes	No, but its absence delays growth.	Presence of a biosynthetic pathway for Val and absence of obvious mutations indicate that lack of growth without this amino acid detected by Hussain <i>et al.</i> (1991) is likely to be an effect of regulation.

### 4.3.3.2 Model analysis for interpretation of experimental results

The *in silico* effect of the removal of amino acids on the objective value (total net flux) and the Glc uptake for production of cell biomass were calculated as described in Section 4.2.3.5. Total net flux through the system was considered as a proxy for the enzymatic cost of the metabolic response. Computing this parameter and the Glc uptake rate for each solution could help explain the experimentally observed effects of nutrient deprivation: if removal of a certain media component allows synthesis of biomass *in silico* but causes a substantial increase in the enzymatic cost, this could potentially be translated into growth being impaired *in vitro*, and the same could be true for instances where the Glc demand increases to levels where the Glc media concentration could become limiting for growth.

The main discrepancy found between the system's behaviour and the experimental results was the *in vitro* lack of growth on MM medium without Pro. The results derived from LP-based analysis of the model (Section 4.2.3.5) were studied in an attempt to identify a metabolic explanation for this and other changes observed *in vitro*. The effect of removing  $\text{NH}_4^+$  or each single amino acid at a time from the medium was calculated as the percentage of increase/decrease in the objective value and the Glc consumption rate with respect to values obtained in standard MM medium (

Table 4-7 and Table 4-8):

**Table 4-7 Summary of the effect caused by removal of single amino acids on the objective value and Glc consumption of the LP-based analysis solutions for biomass production in MM medium when the objective of the analysis was to minimise the total net flux through the network.**

<i>In silico</i> effect caused by removal of single amino acids from the standard MM medium		Objective value (total net flux through the system)	
		Increase	No effect
Glc consumption	Increase	Ala, Asp, His, Glt, Met, Phe, Ser, Trp, Tyr	
	Decrease	Thr	
	No effect	Arg, Asn, Cys, Ile, Leu, Lys, Pro, Val	Gly

**Table 4-8 *In silico* effect of the removal of amino acids and NH<sub>4</sub><sup>+</sup> on the objective value and the Glc uptake rate of responses for production of cell biomass in standard MM medium and modified MM medium when the objective of the analysis was to minimise the total net flux through the network.**

Compound removed	Variation in the objective value (%)	Variation in the Glc uptake (%)
None	-	-
All amino acids and NH <sub>4</sub> <sup>+</sup>	∞	∞
All amino acids	53.1	712
Lys	2.34	0.00
Glt	2.12	3.28
Phe	1.91	14.5
Tyr	1.58	14.5
Cys and Met	1.55	1.53
His	1.43	7.00
Asn	1.10	0.00
Ala	0.98	6.03
Ile	0.915	0.00
Leu	0.801	0.00
Thr	0.624	-2.34
Met	0.615	1.45
Val	0.596	0.00
Trp	0.479	2.19
Arg	0.388	0.00
Asp	0.219	5.62
Ser	0.146	3.14
Cys	0.131	0.00
Pro	0.008	0.00
Gly	0.00	0.00
NH <sub>4</sub> <sup>+</sup>	0.00	0.00

Units: mmol/gDW/h. The variation on the objective value and the Glc uptake rate was calculated as a percentage over the values obtained with the standard *in silico* MM medium (259 mmol/gDW/h = 100% ; and 1.05 mmol Glc (/gDW/h) = 100%).

## 4.4 Discussion

### 4.4.1 Experimental assessment of requirements for amino acids in *S. epidermidis* RP62A

#### Effect of single amino acid deprivation on bacterial growth

Examining the growth curves presented in Section 4.3.1, observations could be made on the effect of media and amino acid deprivation on cell growth:

Cultures grown in MM<sup>-</sup> medium without Ala plateaued at a significantly lower  $A_{600}$  value than those in standard MM medium (Section 4.3.1), indicating that this amino acid is highly important for cell growth.

Cells growing in MM<sup>-</sup> medium lacking Val presented the longest lag phase, lasting between 30 and 48 hours, and the lowest specific growth rate ( $0.10 \text{ h}^{-1}$ ), followed by cultures in MM<sup>-</sup> medium lacking both Cys and Met ( $0.11 \text{ h}^{-1}$ ). This could be indicative of the organism being adapted to utilise these amino acids, and needing time to produce or re-activate the enzymes involved in their synthesis once these compounds are no longer provided with the media. Specifically, the low growth rate exhibited by cultures lacking both Cys and Met is likely due to cells needing to metabolise higher amounts of sulphate in order to synthesise these sulphur-containing amino acids. This would require the participation of a higher number of enzymes to support growth or, at least, an increase in their activity, together with an increase in the activity of the mechanisms mediating the uptake of sulphate. These events might involve upregulating gene expression, which will in turn take some time.

Removal of the amino acids Ala, Arg, Glt, Leu, Thr, Trp and Val caused an apparent delay in growth, while lack of Pro produced an ambiguous result for cell growth. These observations partially match previous findings by Hussain, Hastings and White, who reported a lack of growth or a growth delay in the absence of five of these eight amino acids (Arg, Leu, Thr, Trp and Val) (Hussain *et al.* 1991). The main discrepancies detected between both sets of experimental results correspond to the amino acids Pro and Cys: while these authors reported Pro deprivation to have no effect on cell growth and Cys to be essential for growth, the results obtained in this study were inconclusive on whether RP62A needed external Pro for growth under the conditions tested, while removal of Cys had no apparent effect. It is difficult to conclude if the  $A_{600}$  value exhibited by cultures lacking Pro is due to a very low level of bacterial growth or to other phenomena, such as changes in the chemical components of the medium, an interaction between media components and the bacterial cells or phenotypic changes induced in the cells under these circumstances (e.g. an increase in bacterial adhesion and cell clumping). Either way, if bacterial growth is occurring in these sample, it is greatly

impaired. Thus, the ability of *S. epidermidis* RP62A to grow without Pro was investigated further in the following chapter.

The absence of several amino acid resulted in an increase in the maximum  $A_{600}$  value exhibited by some cultures in comparison with those in standard MM medium. However, this was not associated with an increase in their growth rate or a decrease in their lag phase. For some of them, such as medium samples lacking Thr, this could be explained as the result of nutrient deprivation inducing an increase in the cell adhesion levels, leading to an increase in biofilm formation (Section 4.4.2), however, for others (samples lacking Glt, Leu, His, Phe or Tyr), this was not the case and the reasons behind this effect remain unclear.

#### 4.4.2 Impact of amino acid deprivation on biofilm formation

The results presented in Section 4.3.2 showed how single amino acid deprivation modified the level of biofilm formation in RP62A. However, and despite the fact that biofilm formation in *S. aureus* has been associated with extended growth lag phases and reduced growth rates *in vivo* (Dengler Haunreiter *et al.* 2019), no clear relationship could be established between the effects of amino acid removal observed on cell growth and biofilm formation in this study: for example, by time 48 hours, cultures growing in MM medium lacking certain single amino acids (Glt, Leu, His and Phe) presented  $A_{600}$  values equal or higher than that of those growing in standard MM medium, but exhibited a considerable decrease in their levels of biofilm formation. Specifically, removal of Glt and Leu seemed to have the greatest effect, with respective reductions in biofilm formation of  $94.7\% \pm 5.39$  SD and  $84.3\% \pm 1.92$  SD. This is a strong indication of the importance of these amino acids in the production of biofilm in RP62A, at least, under the conditions considered in this study. On the other hand, removal of other amino acids, such as Ala, caused a substantial decrease in the maximum  $A_{600}$  value but did not lead to a decrease in biofilm formation. Finally, the lack of Arg, Pro and Val caused a decrease in the  $A_{600}$  values measured by time 48 hours while also leading to lower levels of biofilm formation in comparison with those achieved in MM medium. This made hard to differentiate between a real effect in biofilm formation or a lack of production of biofilm components as a direct cause of a reduction in cell growth. Furthermore, since these cultures presented an apparent growth delay and had not reached a growth plateau at the time biofilm formation was measured, it cannot be ruled out that the reduction observed in biofilm production could be a consequence of this phenomenon being delayed due to slow growth, and not necessarily being decreased.

A possible explanation for instances in which nutrient removal led to an apparent reduction in cell growth and an increase in biofilm formation (e.g. absence of Ala) could be that the lack of these components is triggering a stress response leading to the production of EPS (exopolysaccharides,

lipids etc). Bacterial cells are known to produce these compounds under nutrient-restrictive conditions as a way to store them to be utilised as energy sources when needed. Production of EPS increases cell adhesion and cell clumping, which in turn facilitates the formation of biofilms. A phenomenon known as the ‘stringent response’ (Lister *et al.* 2014; Somerville 2016) consisting in the induction of genes linked to antibiotic resistance; expression of phenol-soluble modulins and biofilm formation; activation of genes involved in DNA repair and virulence; and repression of genes related to amino acid metabolism has been described during nutrient deprivation and other adverse conditions, such as heat shock, fatty acid starvation, iron and phosphate limitations (Boutte *et al.* 2013; Somerville 2016). While Leu and Val starvation has been shown to induce the stringent response in *S. aureus* (Geiger *et al.* 2014; Somerville 2016), their removal from MM medium caused an apparent growth delay in RP62A but led to a reduction in biofilm formation. Despite levels of biofilm formation being low in cultures lacking Pro compared to those in cultures in MM medium, these still produced some biofilm biomass (Figure 4-3 ). Thus here we speculate that  $A_{600}$  values of  $0.11 \pm 0.00$  SD detected in these samples could correspond to an increase in the production of EPS in these cell, which could in turn have induced biofilm formation. Planktonic growth was not observed in these cultures.

Changes observed on growth phenotype will surely be linked to a re-wiring of the organism’s metabolic network. It is, however, rather complicated to assign metabolic significance to these results. LP-based analysis specifically applied to the study of energy and biomass production on the biofilm state is described in Chapter 6.

#### **4.4.3 Comparison between experimental data and *in silico* results**

After extensive investigations leading to further refinement of the system, its capability to synthesise amino acids is now reconciled with the genome content of the organism, and, in summary, requirements for essential medium components are either correctly defined by LP-based analysis of the model or can be accounted for in a reasonable manner.

The main discrepancy found between the system’s behaviour *in silico* and the experimental results is the organism’s lack of growth in MM medium without Pro. However, Husain *et al.* did report growth on Pro-deprived cultures. Furthermore, the path for *de novo* synthesis of Pro is missing in RP62A but the presence of genes encoding for a biosynthetic bypass should allow its biosynthesis from ornithine. In the past, authors who investigated growth requirements for staphylococci explained discrepancies between results related to the same organism as a consequence of a delay in the production or activation of vitamin and amino acid synthesising enzymes, which can occur when cells previously cultured on rich media have to adapt to grow in a nutrient-limiting environment

(Gladstone 1937). It is, therefore, plausible that enzymes involved in the biosynthesis of Pro remained inactivated or under-expressed under the conditions tested on this study. Whether the delay in growth observed upon removal of other amino acids is the result of a late induction of the corresponding biosynthesising genes or has other metabolic or regulatory causes is difficult to discern.

As expected, the *in silico* removal of all amino acids from the standard MM medium had a dramatic effect on the objective value (total net flux through the system) and the Glc demand of the LP solutions obtained for biomass production (Table 4-8), increasing the former by 53.1% and the later by 712%, while sole removal of  $\text{NH}_4^+$  did not cause any changes, since  $\text{NH}_4^+$  is not utilised for biomass synthesis in the presence of amino acids. After comparing the outcome of these analyses with results from the *in vitro* growth study (Section 4.3.1) no clear relationship could be established between the *in silico* effect of amino acids removal on the objective value and Glc consumption rate and its *in vitro* effects on bacterial growth and biofilm production:

Although Pro was the only amino acid whose absence led to lack of growth *in vitro* by time 48 hours, its removal from the *in silico* medium produced the smallest increase in the objective value (0.008%) (with the exception of Gly removal, which caused no changes), and had no effect on the rate of Glc uptake.

Removal of Glt and Lys caused the highest impact in the objective value, with a respective increase of 2.12% and 2.34%. While lack of Glt increased the Glc consumption rate by 3.28%, the absence of Lys had no effect on it. However, the *in vitro* effects of their removal were quite different: Glt deprivation caused an apparent delay in cell growth and a substantial decrease in biofilm formation ( $-94.7\% \pm 5.39$  SD), while absence of Lys did not seem to delay growth and reduced biofilm formation by only  $4.43\% \pm 10.64$  SD. This indicates that changes on the total net flux through the system associated with the *in silico* responses on their own are not sufficient to explain alteration on cell growth and biofilm formation.

Removal of the amino acids Phe and Tyr caused the highest effect on the Glc consumption rate, with an increase of 14.5% and caused the objective value to increase between 1.5 and 2% in both cases. However, their corresponding *in vitro* effects were quite different: while lack of Phe or Tyr did not cause an apparent delay in cell growth, Phe removal reduced biofilm formation by  $36.8\% \pm 9.81$  SD, while the absence of Tyr led to a much lower decrease in this parameter ( $-2.92\% \pm 4.56$ ). Therefore, this again indicates that similar *in silico* changes on these parameters cannot be linked to

similar outcomes *in vitro* and are, on their own, insufficient to provide an explanation for the biological responses of the organism.

Finally, investigating the fate of sulphur-containing amino acids and sulphur metabolism in solutions described in Section 4.3.3.2 showed how the model did not take up sulphate as long as Cys was provided. This way, the absence of Met was compensated by synthesising this amino acid from Cys. Biomass production in the absence of Cys involved uptake of Met for its direct incorporation to the biomass and the utilisation of sulphate for Cys synthesis. When neither of them were present, the uptake of sulphate increased in order to allow synthesis of both amino acids. As expected, removal of Cys and Met led to an increase in the total net flux through the system (+1.55%), as well as to an increase in the Glc uptake rate for biomass production (+1.53%). The *in vitro* effect in growth observed in cultures in MM medium without both amino acids was an apparent growth delay and a decrease in the growth rate. However, removal of single amino acids such as Tyr or Phe caused higher increases on these *in silico* parameters without causing an apparent negative effect in cell growth *in vitro*. Therefore, this again exemplifies how these data alone are insufficient to explain the biological changes observed in RP62A.

In summary, thorough investigation of these LP solutions led to the following conclusions:

- i. None of the amino acids are essential for growth *in silico*: removal of each single amino acid at a time from the MM medium does not prevent biomass production.
- ii. Removal of all amino acids *in silico* caused a severe increase on the total net flux through the system and the Glc demand for biomass synthesis. This could potentially impede growth under these conditions and was studied further in the following chapter.
- iii. The lack of effect observed in biomass production upon removal of Pro from the *in silico* MM medium is in conflict with the inconclusive result for growth observed *in vitro*, which was potentially explained as a result of regulation over biosynthesising genes. The small reduction in C and N associated with the removal of single amino acids from the media was considered negligible, since Glc is provided in the medium in sufficient amounts to support cell growth and the presence of 18 other amino acids should suffice to account for the N demand for growth.
- iv. Removal of single amino acids *in silico* can have similar effects on the total net flux and Glc consumption rate of LP solutions for biomass production but can lead to very different outcomes *in vitro*. Study of these parameters on their own is not sufficient to explain the biological behaviour of the organism, which again exemplifies how structural metabolic modelling by itself is insufficient to consider the variability of the complex biological system.

## 4.5 Conclusion

Since, in principle, GSMs represent the full functionality of metabolic networks without considering regulation, predictions derived from their analysis might overestimate the ability of the organisms they represent to grow on nutrient-limited environments. This must be considered when trying to reconcile computational and experimental results.

Certain discrepancies have been observed between the experimental results obtained here and previous phenotypic characterization of RP62A (Hussain *et al.* 1991). Taking into account several findings from authors studying growth requirements for staphylococci (Gladstone 1937; Knight 1937; Emmett *et al.* 1975; Heinemann *et al.* 2005; Lee *et al.* 2009; Bosi *et al.* 2016), a plausible explanation for this could be a delay in the production or activation of synthesizing enzymes as part of an adaptive response to a nutrient-limiting environment dependent on regulation of gene expression. Investigation of changes occurring on the levels of biofilm formation upon nutrient deprivation could not establish a clear relationship between this phenomenon and the changes observed in cell growth. Further work on nutrient utilisation on biofilms is described in Chapter 6. After extensive curation based on genetic analysis and experimental data, the capabilities of the system are in good agreement with the gene content and the *in vitro* behaviour of the organism (Table 4-6). Although the *in vitro* and *in silico* results for growth without Pro seem contradictory, further experimental work described in Chapter 5 showed that RP62A is in fact non-auxotrophic for Pro, which is also in agreement with previous findings by Husain *et al.* (Hussain *et al.* 1991). Hence, it can be concluded that RP62A appears to have no amino acid auxotrophies under the conditions tested and the behaviour of the GSM matches this observation. Computation of the total net flux through the system and Glc consumption rate associated to the *in silico* responses for biomass production upon amino acids removal is, by itself, insufficient to explain the *in vitro* effects observed on cell growth and biofilm formation. This is not surprising, since GSMs generally do not take into account kinetic parameters or regulatory events, and has to be regarded as a current limitation of modelling metabolism at the genome-scale level.

Additional LP-based analysis of the GSM with a focus on N metabolism and amino acids utilisation for biomass production were performed and are described on the following chapter.

## 5 Analysis-guided experimental validation of the GSM on nitrogen metabolism and amino acid utilisation for biomass production

### 5.1 Introduction

Results from Chapter 4 regarding amino acid essentiality were generally in good agreement with the *in silico* results derived from model analysis. Only one clear discrepancy was found: this involved the ability of the system to synthesise biomass in the absence of Pro, for which the corresponding *in vitro* results were inconclusive. This chapter describes additional LP-based analysis of the GSM with a focus on N metabolism and amino acid utilisation for the synthesis of cell biomass (including Pro) and how results derived from these analysis could be used as a guide for experimental design and extended validation of the system.

Specific background relevant to the work described in this chapter on the use of Pro and its metabolism is given below. Other relevant background can be found in Chapter 1, Sections 1.6.1 and 1.6.2.

#### 5.1.1 Investigation of proline essentiality

The *in vitro* dependency on Pro for growth was the main discrepancy identified between the outcome of LP-based analysis for biomass production and the experiments described in Chapter 4. Therefore, this phenomenon was investigated further in this chapter. In order to better understand this matter, relevant literature and metabolic databases were investigated and the findings made are summarised below:

##### 5.1.1.1 Proline biosynthesis

Study of the biosynthetic potential encoded in the genome of RP62A (Chapter 4, Section 4.3.2) showed that it lacks the enzymes involved in Pro biosynthesis from Glt via glutamyl-P but possesses three other enzymes which enable a bypass to obtain this amino acid from ornithine (Lee *et al.* 2009). The sole reaction leading to direct production of Pro in the system is catalysed by the ornithine cyclodeaminase (EC 4.3.1.12). As a result, the GSM was able to reproduce growth in MM<sup>-</sup> medium without Pro, utilising ornithine as a precursor for Pro synthesis. Ornithine is generally obtained from Arg, Gln and Glt. RP62A seems to lack the enzyme arginase (EC 3.5.3.1), which catalyses the hydrolysis of Arg to ornithine and urea in the urea cycle, although it can still produce ornithine from

Arg via citrulline (EC 3.5.3.6 and EC 2.1.3.3). Therefore, while the organism should be able to synthesise ornithine from Gln and Glt, the use of Arg could be partially compromised, thus adding stress to the synthesis of Pro under certain conditions (e.g.: if the availability of Gln and/or Glt is limited, as it is the case in MM medium).

### **5.1.1.2 *The osmoprotective role of proline***

In addition of its well-known proteogenic role, Pro has been described as having an important osmoprotective role in staphylococci growing in high osmolarity environments (Wetzel *et al.* 2011; Somerville 2016). However, the mechanisms through which transport and accumulation of osmolytes allows growth under these conditions is not fully understood yet (Graham *et al.* 1992). Therefore, the upregulation of genes involved in Pro uptake might suggest an adaptation for life in a variety of environments with high salt concentrations (Wetzel *et al.* 2011), including staphylococci preferred niches such as human and animal skin, mucous membranes and blood (Somerville 2016). Since RP62A was originally isolated from the human body, it is possible that, this strain could be adapted to import Pro rather than synthesising it. Therefore, cells could need time to adapt to a Pro-deficient environment, with enzymes involved on Pro anabolism needing to be produced *de novo* or re-activated.

### **5.1.1.3 *Impact of carbon catabolite repression on proline biosynthesis***

Study of *S. aureus* metabolism showed that some of the pathways involved in amino acid catabolism are repressed in the presence of preferred C sources such as Glc (Townsend *et al.* 1996; Li *et al.* 2010; Nuxoll *et al.* 2012; Halsey *et al.* 2017). Results obtained by Halsey *et al.* while studying *S. aureus* growth in HHW medium (Section 4.2.1.1) showed that, in a C rich environment, the catabolism of Arg and Pro and their interconversion was repressed via C catabolite repression (Halsey *et al.* 2017). However, both amino acids were still used for protein synthesis and needed to be supplemented in the medium. When cultures reached their maximum  $A_{600}$  value and Glc was exhausted from the medium, Arg catabolism remained repressed, suggesting that the urea cycle may be subjected not only to C catabolite repression but also to several other layers of regulation. On the other hand, cultures initially grown on HHW medium without Glc efficiently catabolised both amino acids for Glt production, which functioned as the main amino donor for anabolic processes as well as the major C source for growth. In this case, removal of Arg or Pro from the medium did not impair growth. Thus since the Glc concentration of the standard MM medium (0.2% w/v (11.1 mmol)) used in the experiments described in Chapter 4 is similar to that in the HHW medium used by Halsey *et al.* (0.25% w/v (13.9 mmol)) (Halsey *et al.* 2017), C catabolite repression could partly explain the

impairment of growth upon Pro deprivation, and the apparent growth delay observed without Arg (Chapter 4, Section 4.3.1).

### **5.1.2 Use of LP-based analysis to inform experimental design for hypothesis testing**

Taking the information described above into consideration, three main hypotheses can be proposed in order to explain lack of growth without Pro in MM medium: i) the availability of ornithine precursor-amino acids (Glt, Gln and Arg) is limited in MM medium and is insufficient to support Pro biosynthesis; ii) regulation of metabolic genes involved in Pro biosynthesis takes time, and cells still need to adapt to the new nutrient-limited environment, hence growth without Pro is delayed; and/ or iii) since Glc is present in MM medium, the catabolism of Arg and Pro is impaired due to C catabolite repression (Halsey *et al.* 2017).

This chapter describes how LP-based analysis was used for designing the *in vitro* work performed to specifically study the ability of RP62A to: i) grow in the absence of Pro; ii) grow on a set of amino acids (Ala, Arg and Glt) defined as important for biomass production *in silico*; iii) grow on amino acids defined *in silico* as non-suitable N sources for RP62A, and; iv) utilise  $\text{NH}_4^+$  as sole N source for biomass production. The results obtained *in vitro* and *in silico* were compared for further validation of the system.

## **5.2 Methods**

### **5.2.1 Model analysis for the study of the *in silico* utilisation of amino acids for biomass production**

#### ***5.2.1.1 Overall uptake and excretion of amino acids and their uptake to demand ratio for biomass production***

LP-based analysis was applied to the calculation of the  $\text{NH}_4^+$  and amino acids import and export rates during production of cell biomass while satisfying the GAM and NGAM ATP demand (Chapter 3, Section 3.3.4) utilising an *in silico* MM medium. The overall uptake/excretion rates for amino acids during synthesis of planktonic biomass were calculated as the differences in flux between the amino acid medium importers and their biomass exporters. Their uptake to demand ratios were calculated by dividing the flux through each amino acid importer within the flux through their biomass exporter. The *in silico* medium considered in these analyses was MM medium (Chapter 4, Sections 4.2.1.1 and 4.3.1) and the model file was Sepi\_MM.spy.

### ***5.2.1.2 Contribution of individual amino acids to the total nitrogen and carbon uptake for biomass production***

LP-based analysis was used to calculate the individual contribution of each amino acid to the total N and C uptake for production of planktonic biomass while satisfying the GAM and NGAM ATP demand (Chapter 3, Section 3.3.4).

### ***5.2.1.3 Growth on single amino acids as nitrogen sources***

Using LP-based analysis, the system was tested for production of cell biomass while satisfying the growth-associated and non-growth associated ATP demand (Chapter 3, Section 3.3.4). In summary, the *in silico* MM medium was utilised as input and a blockage on the flux through the  $\text{NH}_4^+$  importer was imposed. The import of all amino acids was also blocked and this constraint was then removed for each single amino acid at a time and the analysis repeated in order to identify which of them could be used as single N sources. The effect of utilising single amino acids for biomass production was calculated as the percentage of increase in the objective value with respect of the objective value obtained in the presence of  $\text{NH}_4^+$  and the 19 amino acids.

### ***5.2.1.4 Nitrogen assimilation: glutamate biosynthesis from $\text{NH}_4^+$***

In order to investigate how the system reproduces the organism's capability to assimilate inorganic N from  $\text{NH}_4^+$ , a LP-based analysis was set up with minimisation of total flux as the objective function and generating 1 mmol/gDW/h of Glt as the sole output, with  $\text{NH}_4^+$  and Glc as the only available N and C sources. The solution obtained was investigated and the analysis was repeated upon sequential blocking of flux through key reactions directly involved in Glt production until its synthesis from  $\text{NH}_4^+$  was no longer possible, thus defining a set of reactions that are essential for this process. Note that  $\text{NO}_3^-$  is not present in MM medium but the model could potentially reduce  $\text{NO}_3^-$  to  $\text{NO}_2^-$  and this to  $\text{NH}_4^+$  if needed. The genes associated with these enzymes were investigated in order to determine if some of these reactions could have been incorporated as a result of automatic gap-filling but do not correspond to enzymes encoded by genes truly present in the organism and could be causing an overestimation of the system's capacity to assimilate N. The possible incidence of mutations on these genes in the lab strain was also considered and evaluated following the methods previously described in Chapter 4, Sections 4.3.2 and 4.3.3.

Relevant code to the work described in this chapter can be found in Appendix A, Sections 9.1.3.1 and 9.1.3.2.

## 5.2.2 Experimental design based on the LP-based analysis results

The experimental design for further validation of the model regarding the utilisation of amino acids for biomass production was guided by the results obtained by LP-based analysis on this matter. This work assessed the ability of the organism to:

- i. Grow in the absence of Pro,
- ii. Grow on single amino acids (Ala, Arg and Glt) and on a mixture of them,
- iii. Grow on a mixture of amino acids defined *in silico* as non-suitable N sources for RP62A, and,
- iv. Utilise  $\text{NH}_4^+$  as sole N source.

### 5.2.2.1 Bacterial inoculum

The inoculum utilised for these set of experiments was prepared as previously described in Chapter 4, Section 4.2.1.2.

### 5.2.2.2 Media composition

Note that throughout this document, the term ‘MM medium base’ refers to the composition of the standard MM medium described in Chapter 4, Section 4.2.1.1 without amino acids. The formulation of media batches utilised for growth tests is described below.

- i. Assessment of growth in the absence of Pro:
  - Standard MM medium and MM<sup>-</sup> medium without Pro were prepared as described in Chapter 4, Sections 4.2.1.1 and 4.2.1.3.
  - MM medium base (no amino acids) containing increased concentrations of amino acids which could potentially act as ornithine precursors (Arg and Glt) were prepared as described in point ii.
- ii. Assessment of growth on single amino acids (Ala, Arg and Glt) and on a mixture of them:
  - Modified MM medium batches containing Ala, Arg or Glt or a mixture of the three as sole N source/s were prepared as follows: for simplicity, the total N concentration of the original MM medium (20.4 mmol) was maintained and achieved by supplementation with just one of these amino acids at a time or with a combination of the three as corresponded. This way, the test media were formulated as follows:
    - a. MM medium with Glt as sole N source contained 20.40 mmol of Glt.
    - b. MM medium with Ala as sole N source contained 20.40 mmol of Ala.
    - c. MM with Arg as sole N source contained 5.10 mmol of Arg.
    - d. MM medium with Ala, Arg and Glt contained 7.79 mmol of Ala, 0.341 mmol

of Arg and 11.3 mmol of Glt. This maintained the proportion of N supplied by each amino acid in line with their contribution to the total N uptake for biomass production when available together *in silico*. See LP-based analysis results presented in Section 5.3.2.2 (Table 5-9).

iii. Assessment of growth on a mixture of amino acids defined *in silico* as non-suitable N sources for RP62A:

- A modified test MM medium (tMM) including solely the 7 amino acids that, according to the results derived from model analysis, cannot be utilised N sources in Section 5.3.1.3 was prepared as follows: for simplicity, the concentration of each amino acid present in the tMM medium was made 3.3 times higher than in the original MM medium (Table 5-1). This guaranteed that the total amount of N provided by the mixture was at least as high as in the original medium recipe including all 19 amino acids (20.68 vs 20.37 mmol of N respectively).

**Table 5-1 Concentration of amino acids present in tMM medium and their respective concentration in standard MM medium**

Amino acids identified as non-suitable N sources for BMP	mmol of amino acid in the standard MM medium	mmol of amino acid in the test MM medium (tMM)
Ile	1.14	3.77
Leu	1.14	3.77
Lys	0.68	2.26
Met	0.67	2.21
Phe	0.60	2.00
Trp	0.49	1.62
Tyr	0.55	1.82

iv. Assessment of growth on  $\text{NH}_4^+$  as sole N source:

- For convenience,  $(\text{NH}_4^+)_2\text{HPO}_4$  was used as a supplementary source of  $\text{NH}_4^+$  to be added to the amino acid-free MM medium base. The test samples were formulated as follows:

- a. MM medium base supplemented with 18.7 mmol of  $\text{NH}_4^+$ , thus providing a similar  $\text{NH}_4^+$  concentration to that found in the M9 medium formulation commonly used for staphylococci (Onoue *et al.* 1997; Washburn *et al.* 2001; Wu *et al.* 2012) (Section 5.3.2.4). Here, the Glc concentration of the original MM medium was maintained unmodified at 0.2% w/v (11.10 mmol).
- b. MM medium supplemented with 18.7 mmol of  $\text{NH}_4^+$  and double Glc

concentration (0.4% w/v or 22.10 mmol) in order to account for a hypothetical dramatic increase on Glc demand if biofilm formation is to be triggered by the stringent response (Lister *et al.* 2014; Somerville 2016).

- c. MM medium supplemented with 60.6 mmol of  $\text{NH}_4^+$ , thus providing a similar  $\text{NH}_4^+$  concentration to that in the AAM medium (Rudin *et al.* 1974) for staphylococci (Section 5.3.2.4), and increasing the original Glc concentration 10 times (2% w/v or 111 mmol) for the same reason as above.

### 5.2.2.3 Experimental setup

Cell inocula were obtained as described above and added in a proportion 1/100 to 15 ml sterile tubes containing 10 ml of the corresponding test medium, given an approximate final cell concentration of  $1.5 \cdot 10^6$  CFU/ml per sample. The tubes were incubated at 37 °C and 180 rpm for a period of 7 days. Cell growth was monitored by measuring absorbance spectrophotometrically at 600 nm ( $A_{600}$ ) every 24 hours. Standard MM medium samples were inoculated and used as a positive control for growth. Non-inoculated standard MM medium and test media samples were used as sterility controls for each growth condition. All samples, including controls, were prepared as three independent biological replicates and the growth data obtained were expressed as the means from these replicates. Three extra replicates were prepared and incubated unopened. For these, growth presence or absence was monitored by visual inspection, thus reducing risk of obtaining a positive result for growth due to contamination to a minimum. Background levels were calculated by averaging values obtained for the sterility controls and were subtracted from the absorbance measurements of the tests samples as corresponded. At the end of the incubation period, those samples where growth was detected were streaked out on BHI agar in order to discard a possible contamination. All tubes were incubated for an extra period of 7 days and monitored again for growth by visual inspection after a total of 14 days of incubation.

## 5.3 Results

### 5.3.1 *In silico* utilisation of amino acids for biomass production

These analysis investigated which amino acids were preferentially utilised by the system for biomass production. The results obtained showed which of them were taken up at a higher proportion than needed for their direct incorporation into the biomass. The proportion in which each amino acid contributed to the total N and C uptake for biomass synthesis was also identified. Furthermore, lists of amino acids which could and could not be utilised as sole N sources for growth were obtained, and, finally, the system was analysed for optimal responses for Glt synthesis from  $\text{NH}_4^+$ . This work

expanded our understanding on the capabilities of the cell to metabolise amino acids and provided useful data to design experiments for further validation.

### ***5.3.1.1 Overall uptake and excretion of amino acids and their uptake to demand ratio for biomass production***

Calculation of the total uptake or excretion of amino acids and  $\text{NH}_4^+$  and their uptake to demand ratios during planktonic growth was performed as described in Section 5.2.1.1. The results obtained are summarised in Table 5-2:

**Table 5-2 Overall uptake or excretion of  $\text{NH}_4^+$  and amino acids during *in silico* production of planktonic biomass in standard MM medium and their uptake to demand ratios in the presence of  $\text{O}_2$ .**

<b>Compound</b>	<b>Overall uptake/excretion (mmol/gDW/h)</b>	<b>Ratio uptake to demand</b>
<b>Glt</b>	9.29	47.6
<b>Ala</b>	1.75	9.26
<b>Thr</b>	0.857	5.79
<b>Asp</b>	0.366	2.40
<b>Arg</b>	0.15	2.36
<b>Lys</b>	0.103	1.44
<b>Met</b>	0.007	1.09
<b>Asn</b>	0.00	1.00
<b>Cys</b>	0.00	1.00
<b>His</b>	0.00	1.00
<b>Ile</b>	0.00	1.00
<b>Leu</b>	0.00	1.00
<b>Phe</b>	0.00	1.00
<b>Pro</b>	0.00	1.00
<b>Ser</b>	0.00	1.00
<b>Trp</b>	0.00	1.00
<b>Tyr</b>	0.00	1.00
<b>Val</b>	0.00	1.00
<b>Gly</b>	-0.190	0.00
<b>Gln</b>	-0.200	0.00
<b><math>\text{NH}_4^+</math></b>	-7.87	-

A positive overall uptake/excretion value indicates compound net uptake, a negative value indicates net excretion and a value of 0 indicates that the compound was taken up at the same proportion as it was excreted in the biomass. An uptake to demand ratio value of 0 indicates lack of uptake, a value of 1 indicates that the

amino acid was taken up at the same proportion as it was excreted, while a value  $> 1$  indicates that the amino acid was taken up at a higher proportion than it was incorporated into the biomass.

The amino acids Asn, Cys, His, Ile, Leu, Phe, Pro, Ser, Trp, Tyr and Val presented uptake to demand ratios of 1 and, therefore, were directly incorporated to the protein in the biomass. Amino acids with ratios of 0 (Gln and Gly) were not taken up from the medium, but rather produced by the network, and, finally, those amino acids with uptake to demand ratios  $> 1$  (Glt, Ala, Thr, Asp, Arg, Lys and Met) were imported and used for anabolic purposes.

For completeness, since Cys and Met are the only sulphur containing amino acids, their fate during biomass production was investigated further: while Cys was taken up in the same proportion as it was incorporated into the biomass, Met was taken up at a slightly higher rate than it was excreted as protein (0.007 mmol/gDW/h). Study of the reactions involved in the LP solution showed that Met was being utilised to obtain homo-Ser via homo-Cys, producing 0.007 mmol/gDW/h of hydrogen sulphide ( $H_2S$ ) in the process, which was then excreted in order to balance the atoms of sulphur in the system. Finally, homo-Ser was used to produce Asp in a NAD and NADP reducing process which generate 0.007 mmol/gDW/h of NADH and NADPH respectively. Reduction of these co-substrates is advantageous, since they could then be utilised to either obtain ATP via the ETC (NADH) or in biosynthetic processes (NADPH).

### ***5.3.1.2 Contribution of individual amino acids to the total nitrogen and carbon uptake for biomass production***

This section focuses on understanding how the uptake of different amino acids contributes to the total N and C uptake for synthesis of planktonic biomass. The percentage in which each amino acid contributed to the N demand was calculated as described in Section 5.2.1.2 and the results obtained are summarised in Table 5-3:

**Table 5-3 Contribution of each amino acid to the total N taken up for synthesis of planktonic biomass when  $\text{NH}_4^+$  and all 19 amino acids were present in the standard MM medium in the presence of  $\text{O}_2$  and total net flux through the network was minimised.**

<b>Amino acid</b>	<b>Contribution to the total N uptake (%)</b>
<b>Glt</b>	56.2
<b>Ala</b>	11.6
<b>Arg</b>	6.19
<b>Thr</b>	6.13
<b>Lys</b>	4.00
<b>Asp</b>	3.71
<b>Asn</b>	2.04
<b>Leu</b>	1.67
<b>Ile</b>	1.59
<b>His</b>	1.30
<b>Val</b>	1.23
<b>Ser</b>	1.17
<b>Phe</b>	0.813
<b>Tyr</b>	0.703
<b>Pro</b>	0.688
<b>Met</b>	0.539
<b>Trp</b>	0.272
<b>Cys</b>	0.114
<b>Gly</b>	0.00
<b>Total</b>	100

For completeness, the uptake of N as either  $\text{NH}_4^+$  or amino acids when all amino acids were present or absent in the medium was investigated, as well as the excretion of N as different compounds under these conditions (Table 5-4):

**Table 5-4 Uptake and excretion of N for the production of planktonic biomass when amino acids were either absent (i) or present (ii) in MM medium in the presence of O<sub>2</sub> while total net flux through the network was minimised**

	Without amino acids	With amino acids
<b>N imported and exported during BMP</b>	mmol/gDW/h	mmol/gDW/h
<b>N taken up as NH<sub>4</sub><sup>+</sup> from the medium</b>	+ 7.25	0.00
<b>N taken up as amino acids from the medium</b>	0.00	+ 16.9
<b>N exported as amino acids in the biomass</b>	- 4.40	- 4.40
<b>N excreted as by-products (NH<sub>4</sub><sup>+</sup>)</b>	0.00	- 9.64
<b>N exported as other biomass components</b>	- 2.85	- 2.85
<b>N taken up minus N excreted</b>	0.00	0.00

These data show how Glt contributed to the total N uptake in a significantly higher proportion than any other amino acid, accounting for approximately 56% of the total N demand, followed by Ala which accounted for approximately a 12%. The rest of the amino acids contributed to this in the following order (from higher to lower): Arg, Thr, Lys, Asp, Asn, Leu, Ile, His, Val, Ser, Phe, Tyr, Pro, Met, Trp and Cys.

The total N demand in the absence of amino acids was calculated by removing them from the *in silico* MM medium and re-solving the LP-based analysis for biomass production with NH<sub>4</sub><sup>+</sup> as sole N source. Under these conditions, 7.25 mmol/gDW/h of NH<sub>4</sub><sup>+</sup>, and therefore 7.25 mmol/gDW/h of N, were taken up and fully consumed during biomass synthesis, with no N-containing by-products being excreted. When NH<sub>4</sub><sup>+</sup> and all amino acids were available, the total N uptake increased up to 16.9 mmol/gDW/h, with all N being taken up as amino acids and 9.64 mmol/gDW/h being excreted as NH<sub>4</sub><sup>+</sup> as a result of amino acid catabolism. In both cases, 4.40 mmol/gDW/h of N were excreted as biomass protein and 2.85 mmol/gDW/h as other biomass products.

The utilisation of amino acids as C sources to generate energy and biomass was studied as a way to understand the link between energy production and the system's biosynthetic network. The percentage in which each amino acid contributed to the C demand for planktonic growth was calculated as described in Section 5.2.1.2 and the results obtained are summarised in Table 5-5:

**Table 5-5 Contribution of each amino acid to the total C uptake for synthesis of planktonic biomass when  $\text{NH}_4^+$  and all 19 amino acids were present in the standard MM medium in the presence of  $\text{O}_2$  when total net flux through the network was minimised.**

<b>Amino acid</b>	<b>Contribution to the total C uptake (%)</b>
<b>Glt</b>	59.6
<b>Ala</b>	7.41
<b>Thr</b>	5.20
<b>Asp</b>	3.15
<b>Lys</b>	2.55
<b>Leu</b>	2.12
<b>Ile</b>	2.03
<b>Arg</b>	1.97
<b>Phe</b>	1.55
<b>Tyr</b>	1.34
<b>Val</b>	1.30
<b>Asn</b>	0.868
<b>Ser</b>	0.745
<b>Pro</b>	0.73
<b>Met</b>	0.572
<b>His</b>	0.553
<b>Trp</b>	0.317
<b>Cys</b>	0.072
<b>Gly</b>	0.00
<b>Total</b>	100

For completeness, the difference in the C obtained from Glc in the presence and absence of amino acids was investigated. For that, the same LP-based analysis was repeated, now blocking flux through all amino acid importers. The uptake and excretion of C as different compounds is compared for both solutions in Table 5-6:

**Table 5-6 Uptake and excretion of C for the production of planktonic biomass when amino acids were either absent (i) or present (ii) in MM medium in the presence of O<sub>2</sub> when total net flux through the network was minimised.**

	<b>Without amino acids</b>	<b>With amino acids</b>
<b>C imported and exported during BMP</b>	<b>mmol/gDW/h</b>	<b>mmol/gDW/h</b>
<b>C taken up as Glc</b>	+51.03	+6.28
<b>C taken up as amino acids form the medium</b>	0.00	+73.3
<b>C exported as amino acids in the biomass</b>	-16.5	-16.5
<b>C excreted as by-products</b>	-18.9	- 47.5
<b>C exported as other biomass components</b>	-15.6	-15.6
<b>C taken up minus C excreted</b>	0.00	0.00

The C uptake as Glc when this is the only available C source (i) was 51.03 mmol/gDW/h. From this, 18.9 mmol/gDW/h were excreted as metabolic by-products (Ac, CO<sub>2</sub> and Form). When Glc and all 19 amino acids in the standard MM medium were available (ii), the import of C as Glc fell drastically, although some Glc was still utilised (accounting for the 7.89 % of the total C demand), while the import of C as amino acids (73.3 mmol/gDW/h) was higher than the import of C as Glc in (i). In consequence, the total C uptake of the process was much higher under these circumstances (+55.9 %), and, therefore, so was the excretion of C as metabolic by-products (CO<sub>2</sub> and Suc), maintaining the total amount of C uptaken equal to the amount of C excreted. In both cases, 16.5 mmol/ gDW/h of C were excreted as biomass protein and 15.6 mmol/gDW/h as other biomass components.

### ***5.3.1.3 Growth on single amino acids as nitrogen sources***

As already seen in this chapter, when amino acids are available, NH<sub>4</sub><sup>+</sup> is not essential for growth. However, not all amino acids can be utilised by the system as sole N sources. Defining which single amino acids can and cannot be used for biomass production help us to complete our understanding of the cell metabolic capabilities, while providing useful data to design experimental assays for validation purposes. Thus each single amino acid was tested for its suitability as sole N source for growth support. The impact on the total net flux of the LP solution of growing in a single amino acid rather than in standard MM medium was calculated as described in Section 5.2.1.3. The results obtained are summarised in Table 5-7.

**Table 5-7 Variation on the total net flux of the LP solution when just a single N source was available at a time for production of planktonic biomass in the presence of O<sub>2</sub> when total net flux through the network was minimised.**

<b>N source available</b>	<b>Objective value (mmol/gDW/h)</b>	<b>Variation (%)</b>
<b>Met</b>	3468	1237
<b>NH<sub>4</sub><sup>+</sup></b>	397	53.1
<b>Cys</b>	384	48.3
<b>Gly</b>	376	45.0
<b>Ser</b>	369	42.2
<b>His</b>	365	40.8
<b>Asn</b>	364	40.3
<b>Asp</b>	362	39.6
<b>Thr</b>	350	35.1
<b>Val</b>	350	34.8
<b>Ala</b>	335	29.2
<b>Arg</b>	329	26.8
<b>Pro</b>	327	26.1
<b>Glt</b>	320	23.9
<b>NH<sub>4</sub><sup>+</sup> and all amino acids</b>	259	-

The effect on the objective value (total net flux) was calculated with respect to the value obtained in the standard MM medium: 259 mmol/gDW/h = 100%.

The results obtained showed that 13 of the 19 amino acids in MM medium could be utilised as single N sources for biomass production (Table 5-7), with those being: Ala, Arg, Asn, Asp, Cys, Glt, Gly, His, Met, Pro, Ser, Thr and Val. The remaining amino acids could not be used for this purpose when available either independently or in combination.

Inspection of the organism-specific BioCyc and KEGG databases and the RP62A genome sequence showed that the system lacks the enzymatic machinery needed to fully catabolise Ile, Leu, Lys, Phe, Trp and Tyr, which explains why they could not support growth on their own. A complete pathway for the utilisation of Met is present, and all biomass components could be individually produced with Met as sole N source. However, LP-based analysis results showed that Met catabolism supported *in silico* biomass production in the absence of other N sources as long as the upper flux bounds for exporters of biomass components were relaxed and the system was allowed to export menaquinones at a higher rate than that required for the production of cell biomass. The catabolism of Met as sole

N source also led to higher production and excretion rates of autoinducer-2 and H<sub>2</sub>S than those identified for growth in standard MM medium. Furthermore, this response was associated with a dramatic increase in the total net flux through the system (+1,237%) and the Glc uptake (+10,482%) when compared to growth in standard conditions.

#### **5.3.1.4 Nitrogen assimilation: glutamate biosynthesis from NH<sub>4</sub><sup>+</sup>**

The ability of the system to utilise NH<sub>4</sub><sup>+</sup> for Glt synthesis was investigated by applying the methods described in Section 5.2.1.4. This allowed us to identify which reactions are involved in the process, define their essentiality and compare these responses to bacterial biological strategies for the assimilation of N: initially, an optimal LP-based analysis solution was obtained, involving 64 reactions and allowing synthesis of 1 mmol/gDW/h of Glt while consuming 1 mmol/gDW/h of NH<sub>4</sub><sup>+</sup> and Glc and excreting CO<sub>2</sub> and Form. In this solution, the system preferentially used two Glt dehydrogenases (Glt-dhs). These Glt-dhs were a NAD-dependent (EC 1.4.1.2) and a NADP-dependent enzyme (EC 1.4.1.4), and generated Glt from NH<sub>4</sub><sup>+</sup> and 2-KG, while respectively oxidising NADH and NADPH. These enzymes worked in combination with the Gln synthetase or GS (EC 6.3.1.2), producing Gln from Glt while consuming ATP and NH<sub>4</sub><sup>+</sup>, and an aminotransferase reaction (EC 4.3.2.10) from the His-biosynthesis pathway, regenerating Glt from Gln while producing intermediates for the synthesis of His. This formed a cycle of production and consumption between Glt and Gln. Re-solving the LP-based analysis while blocking each of the Glt-dhs at a time showed that they can work together or independently. When both Glt-dhs were blocked, a solution was obtained, involving the Gln synthase / Glt synthase (GS/GOGAT) cycle: this included both, the NADH-dependent Glt synthase or GOGAT (EC 1.4.1.14), which produced two molecules of Glt from Gln and 2-KG, and the GS (EC 6.3.1.2), which again regenerated Gln from Glt. When flux through the GOGAT reaction was blocked, a new solution was obtained, where this enzyme was replaced by its NADPH-dependent equivalent (EC 1.4.1.13). Upon blockage of flux through the EC 1.4.1.13 reaction, one last feasible solution was obtained, where a ferredoxin-dependent Glt synthase (EC 1.4.7.1) took over the production of Glt. If flux was still allowed through the GOGAT reactions (EC 1.4.1.14/3) but the GS reaction was blocked, a final feasible solution was generated, involving the NADH-dependent GOGAT, and a GlcN-F6P aminotransferase (EC 5.3.1.19), which allowed conversion of Glt back to Gln while transforming GlcN-6P onto F6P. This, again, formed a cycle of production and consumption between Glt and Gln. Finally, blocking flux through this last aminotransferase reaction also prevented Glt synthesis.

The presence or absence in the genome of RP62A of genes encoding the enzymes involved in the solutions described above was investigated. There were no genes found encoding for the ferredoxin-dependent Glt synthase (EC 1.4.7.1), thus its corresponding reaction was removed from the model.

### **5.3.2 Experimental work for model validation on amino acid utilisation for biomass production**

This section exemplifies how model analysis can be used to orientate experimental design in order to investigate an hypothesis of interest and, in this case, to perform further validation of the model. For this purpose, a set of experiments were designed based on the LP-based analysis results described above. These assessed the ability of the organism to:

- i. Grow in the absence of Pro,
- ii. Grow on a set of key single amino acids (Ala, Arg and Glt) and on a mixture of them,
- iii. Grow on a mixture of amino acids defined *in silico* as non-suitable single N sources, and,
- iv. Utilise  $\text{NH}_4^+$  as sole N source.

#### ***5.3.2.1 Assessment of growth in the absence of proline***

In order to finally clarify whether or not Pro is essential for growth in RP62A, a final set of experiments was proposed based on the following criteria:

i. Adaptation to grow without Pro could take some time, since enzymes involved in its biosynthesis could either have to be re-activated or synthesised *de novo* via gene regulation (Section 5.1.1.2). Therefore, cell growth in MM<sup>-</sup> medium without Pro was tested again, but this time monitored for longer than 48 hours, which was the incubation time of the growth experiments described in Chapter 4, Section 4.2.1.

ii. It is possible that the amount of amino acids in the standard MM medium which could act as ornithine precursors (Arg and Glt) is not sufficient to support Pro synthesis when this amino acid is missing (Section 5.1.1.1). Therefore, their concentration was increased in the media formulated to test this hypothesis (Section 5.2.2.2).

#### ***5.3.2.2 Assessment of growth on single amino acids (alanine, arginine and glutamate) and on a mixture of them***

Considering the results obtained from the LP-based analysis described in Sections 5.3.1.2, and 5.3.1.3, a set of amino acids could be selected for testing the ability of the organism to utilise them as sole N sources for growth *in vitro* for further validation purposes: within those amino acids that could act as single N sources, the three that contributed in a higher proportion towards the total N uptake for biomass production (Section 5.3.1.2, Table 5-3) were Glt (56.2%), Ala (11.6%) and Arg

(6.19%). Furthermore, when available individually, these amino acids yielded solutions with the lowest objective values, excepting Pro (Section 5.3.1.3, Table 5-7), suggesting that they could be potentially utilised for biomass production by the organism with the lowest associated enzymatic costs. These LP-based analysis solutions were investigated further and the data summarised below was used to design experiments to assess if growth on these amino acids is possible when provided independently or in combination (Table 5-8 and Table 5-9):

**Table 5-8 Main features of the LP-based analysis solutions obtained when Ala, Arg and Glt were available as sole N sources for biomass production in MM medium in the presence of O<sub>2</sub> when total net flux through the network was minimised.**

Single N source available	Reactions in solution (n)	Objective value	Amino acid uptake	Glc uptake	NH <sub>4</sub> <sup>+</sup> export	By-products exported
Ala	341	335	15.6	2.04	-8.39	Ac, CO <sub>2</sub> , Form
Arg	328	329	4.89	3.19	-12.3	CO <sub>2</sub> , Form
Glt	330	320	9.34	2.21	-2.09	Ac, CO <sub>2</sub> , Form, Suc

Units: mmol/gDW/h

**Table 5-9 Main features of the LP-based analysis solution obtained when a mixture of Ala, Arg and Glt was available as sole N source for biomass production in MM medium in the presence of O<sub>2</sub> when total net flux through the network was minimised.**

N sources available	Reactions in solution (n)	Objective value	Amino acid uptake	Contribution to total N uptake (%)	Glc uptake	NH <sub>4</sub> <sup>+</sup> export	By-products exported
Ala Arg Glt	328	310	5.96 0.261 8.62	38.1 6.68 55.2	2.09	-8.37	Ac, CO <sub>2</sub> , Form, Suc

Units: mmol/gDW/h

### ***5.3.2.3 Assessment of growth on a mixture of amino acids defined as non-suitable nitrogen sources for RP62A***

According to the LP-based analysis results described in Section 5.2.1.3, from the 19 amino acids present in MM medium, 6 amino acids (Ile, Leu, Lys, Phe, Trp and Tyr) cannot be utilised by the system as N sources to support growth when available either individually or in combination. Utilisation of Met is theoretically possible, but caused a big increase in the total net flux through the

system (+1,237%) and the Glc demand (+10,482%) when compared to growth in standard MM medium. This effect was far more dramatic than that observed on these parameters during growth on any other single amino acid or on  $\text{NH}_4^+$  as sole N source. Therefore, based on these results, Met was also considered as a non-suitable N source for RP62A during the design of this experimental work.

#### **5.3.2.4 Assessment of growth on $\text{NH}_4^+$ as sole nitrogen source**

The literature regarding minimal media without amino acids utilised to grow staphylococci was investigated: this showed that M9 medium (Onoue *et al.* 1997; Washburn *et al.* 2001; Wu *et al.* 2012) has been vastly used for this purpose and contains  $\text{NH}_4^+$  as  $\text{NH}_4^+\text{Cl}$  at a concentration of 18.69 mmol and 22.2 mmol of Glc. In 1974, Rudin *et al.* utilised a chemically defined minimal medium (AAM) (Rudin *et al.* 1974) specifically developed for the isolation of amino acid-requiring mutants of *S. aureus*: in AAM 60.53 mmol of  $\text{NH}_4^+$  are provided as  $(\text{NH}_4^+)_2\text{SO}_4$  while the concentration of Glc is 27.69 mmol. The standard MM medium utilised throughout this project contains as little as 0.03 mmol of  $\text{NH}_4^+$  (provided as  $(\text{NH}_4^+)_2\text{SO}_4 \cdot \text{FeSO}_4 \cdot 6\text{H}_2\text{O}$ ), and a Glc concentration of 11.1 mmol. This means that the amount of  $\text{NH}_4^+$  provided in our medium will not suffice to account for the N demand for growth on its own. Therefore, this was taken into account when formulating the test media samples for *in vitro* evaluation of the ability of RP62A to grow utilising either  $\text{NH}_4^+$  as sole N source, single amino acids or a combination of just three amino acids as N sources, as explained in Section 5.2.2.2.

It is important to keep in mind that removing amino acids from the media might trigger physiological changes in the cells, such as inducing growth in the form of a biofilm and/or repression of amino acid metabolism due to the stringent response (Lister *et al.* 2014; Somerville 2016) which might in turn have unpredicted biological effects (e.g. an increase on the Glc demand). Furthermore, other studies have shown that biofilm production is significantly enhanced at high Glc concentrations (1%, 2.7%, 5%, an even 10% (w/v)) (Lim *et al.* 2004; Agarwal *et al.* 2013; Fernanda Cristina Possamai Rossatto 2017), which again suggest that if a phenotypic change of this nature takes place, the Glc demand might increase significantly.

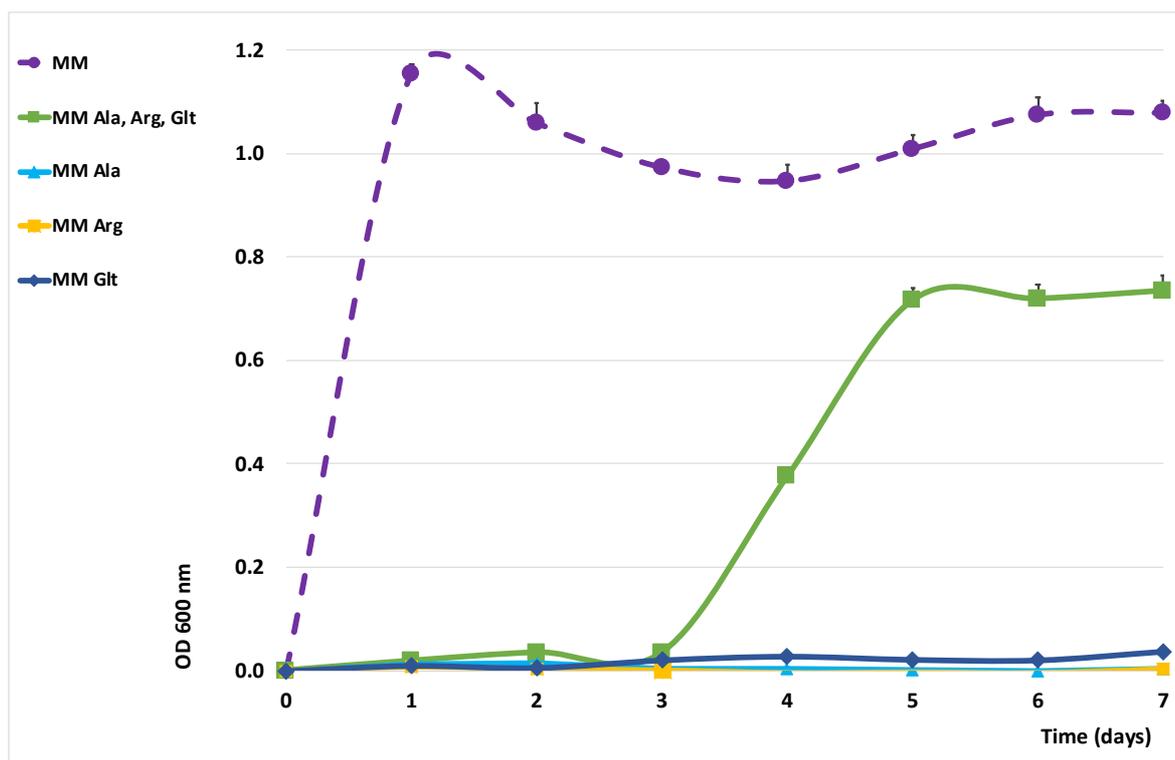
#### **5.3.2.5 Experimental results**

Taking into account the data summarised in Sections 5.3.2.1 to 5.3.2.4, an *in vitro* study was designed in which RP62A was monitored for growth in standard MM medium vs:

- i. MM<sup>-</sup> medium without Pro
- ii. MM medium base supplemented with either Ala, Arg or Glt or with a mixture of them,
- iii. MM medium base supplemented with a mixture of amino acids non-suitable as N sources

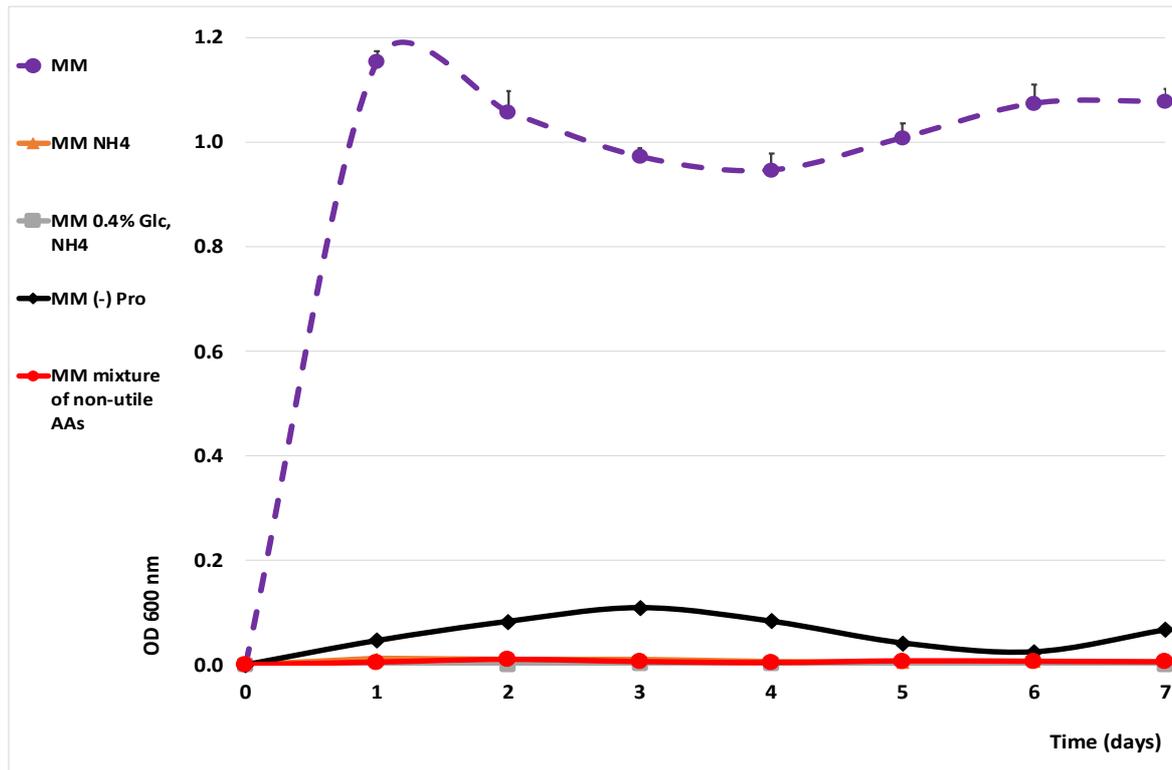
iv. MM medium base supplemented with a range of Glc and  $\text{NH}_4^+$  concentrations

For further details on the composition of these media and the experimental set up see Section 5.2.2.2. The results obtained were used to construct growth curves for RP62A in these media (Figure 5-1 and Figure 5-2) and are summarised in Table 5-10. For clarity, since no growth was detected on any of the sterility controls neither on the samples containing the highest amount of Glc and  $\text{NH}_4^+$ , their growth curves were not included in these figures.



**Figure 5-1 Growth curves for *S. epidermidis* RP62A cultures in modified MM medium samples in comparison to cultures in standard MM medium over a period of 7 days**

Legend: purple dashed line (MM) = standard MM medium; green line (MM Ala, Arg, Glt) = MM medium base supplemented with a mixture of Ala, Arg and Glt; light blue line (MM Ala) = MM medium base supplemented with Ala; yellow line (MM Arg) = MM medium base supplemented with Arg; dark blue line (MM Glt) = MM medium base supplemented with Glt. Each data point corresponds to the mean  $A_{600}$  value of three independent biological replicates. Error bars = SEM.



**Figure 5-2 Growth curves for *S. epidermidis* RP62A cultures in modified MM medium samples in comparison to cultures in standard MM medium over a period of 7 days**

Legend: purple dashed line (MM) = standard MM medium; orange line (MM NH<sub>4</sub><sup>+</sup>) = MM medium base supplemented with 15.1 mmol of NH<sub>4</sub><sup>+</sup>; grey line (MM 0.4% Glc, NH<sub>4</sub><sup>+</sup>) = MM medium base supplemented with 15.1 mmol of NH<sub>4</sub><sup>+</sup> and double the standard Glc concentration (22.1 mmol, 0.4% w/v); black line (MM (-) Pro) = MM<sup>-</sup> medium without Pro; red line (MM mixture of AAs defined as non-suitable N sources) = MM medium base containing no NH<sub>4</sub><sup>+</sup> and supplemented with a mixture of the amino acids defined as non-suitable N sources in Section 5.3.1.3. Each data point corresponds to the mean *A*<sub>600</sub> value of three independent biological replicates. Error bars = SEM.

**Table 5-10  $A_{600}$  values for *S. epidermidis* RP62A cultures in modified MM medium containing different N sources after 7 days of incubation**

<b>Time: 7 days</b>	
<b>N source/s in the medium</b>	<b><math>A_{600}</math></b>
All amino acids in MM medium	1.08 ± 0.04
Ala, Arg and Glt	0.73 ± 0.05
Ala	0.00 ± 0.00
Arg	0.00 ± 0.01
Glt	0.00 ± 0.00
Amino acids non-suitable as N sources	0.00 ± 0.00
(NH <sub>4</sub> <sup>+</sup> ) <sub>2</sub> HPO <sub>4</sub>	0.00 ± 0.00

Orange colour indicates that growth was observed *in vitro* while blue colour indicates lack of growth. Error values = SD.

For comparison purposes, the  $A_{600}$  values for cultures in standard MM medium and MM<sup>-</sup> medium without Pro at time points 2, 3 and 7 days of incubation are shown below (Table 5-11):

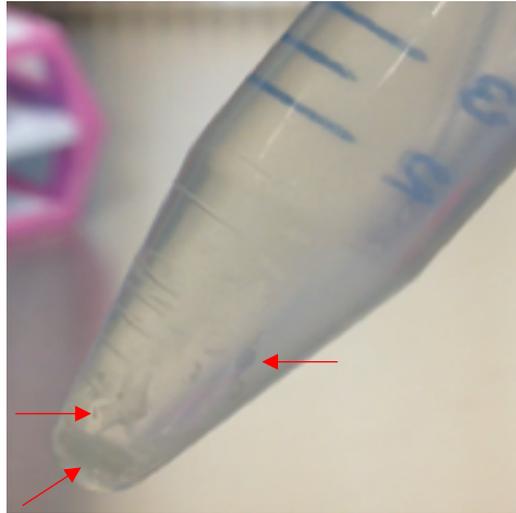
**Table 5-11  $A_{600}$  values for *S. epidermidis* RP62A cultures in MM medium and MM<sup>-</sup> medium without Pro after 2 and 7 days of incubation**

	<b>MM medium</b>	<b>MM<sup>-</sup> medium without Pro</b>
<b><math>A_{600}</math> Day 2</b>	1.06 ± 0.07	0.08 ± 0.00
<b><math>A_{600}</math> Day 3</b>	0.97 ± 0.03	0.11 ± 0.00
<b><math>A_{600}</math> Day 7</b>	1.08 ± 0.04	0.07 ± 0.01

Error values = SD

Interestingly, cells incubated in the absence of Pro reached a maximum  $A_{600}$  value of 0.11 ± 0.00 SD, which matched the maximum  $A_{600}$  value observed for samples incubated under the same conditions in previous experiments (Chapter 4). This, together with the fact that cultures in MM<sup>-</sup> medium without Pro presented a visible amount of biofilm biomass attached to the walls of the incubation tubes (Image 5-1) is in line with the hypothesis proposed in Chapter 4 that attributed this  $A_{600}$  value to cells undergoing a phenotypic change leading to an increase in their adhesive properties, causing higher levels of intercellular attachment and cell clumping, as is commonly observed during biofilm formation. Therefore, these results seem to indicate that cells cultured in standard MM<sup>-</sup> medium

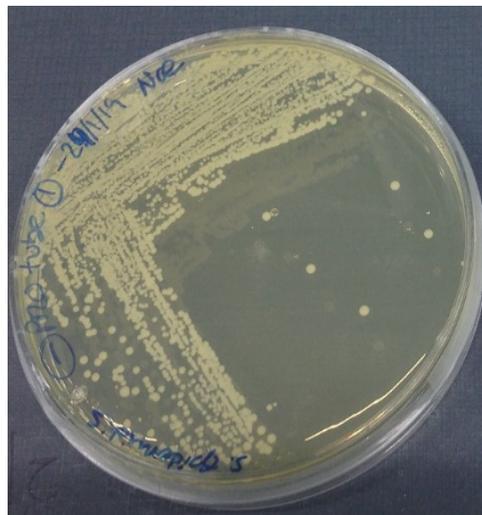
without Pro do not grow but rather form biofilms, possibly, doing so as a response to the stress induced by nutrient deprivation.



**Image 5-1 Tube containing *S. epidermidis* RP62A in MM<sup>-</sup> medium without Pro after 2 days of incubation**

Biofilm formation observed on the bottom and the walls of the tube as white slimy material is pointed out by red arrows.

As Image 5-2 shows, *S. epidermidis* cells were still present in these cultures, and no contamination was detected by eye on the plates.

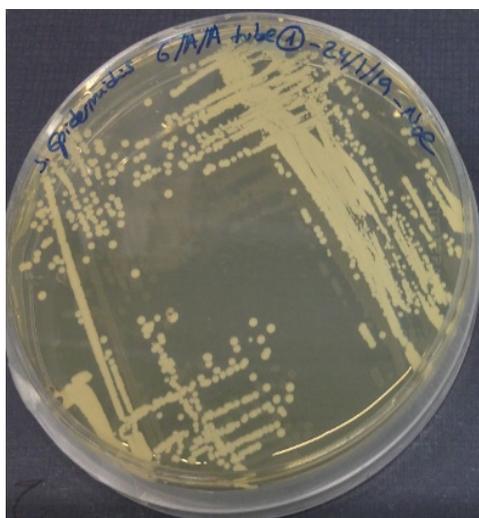


**Image 5-2 BHI agar plate streaked out with a *S. epidermidis* RP62A culture in MM<sup>-</sup> medium without Pro after 7 days of incubation**

This plate was incubated at 37°C for 24 hours and inspected by eye in order to discard contamination.

The only two media formulations that seemed to support growth *in vitro* were the standard MM medium (positive growth control) and the MM medium base containing a mixture of Ala, Arg and Glt. Cultures in standard MM medium presented an average  $A_{600}$  value of  $1.15 \pm 0.03$  SD by day 1, when they seemed to have reached stationary phase, with values remaining around 1 after 7 days of incubation ( $1.08 \pm 0.04$  SD). In contrast, cultures growing in MM medium base with the three amino acid mixture only started to show growth after 3 days of incubation, and seemed to have reached stationary phase by day 5 ( $0.72 \pm 0.04$  SD) and continued exhibiting a similar  $A_{600}$  value by day 7 ( $A_{600} = 0.73 \pm 0.05$  SD). All other samples presented  $A_{600}$  values below 0.0 throughout the whole experiment. No apparent changes on the turbidity of the cultures were detected by eye after 14 days of incubation, thus confirming that media that did not support growth by 7 days still did not support growth after doubling the incubation time. Furthermore, those samples that showed growth did also showed growth on their un-opened replicates, which helped to rule out a possible contamination of the samples monitored spectrophotometrically.

Homogeneous *S. epidermidis* colonies could be seen on plates streaked out with cultures grown in MM medium base and a combination of Ala, Arg and Glt (Image 5-3), with no contaminants detected by eye, thus ruling out a possible contamination.



**Image 5-3 BHI agar plate streaked out with a *S. epidermidis* RP62A culture in MM medium base containing a combination of Ala, Arg and Glt after 7 days of incubation**

This plate was incubated at  $37^{\circ}\text{C}$  for 24 hours and inspected by eye in order to discard contamination.

The changes in the LP-based analysis solution for growth when the amino acid composition of the standard MM medium was modified were investigated, and the effects on the total net flux through

the system and the Glc uptake were compared in an attempt to identify possible explanations for the growth effects observed *in vitro* (Table 5-12):

**Table 5-12 Summary of the *in silico* effect of amino acid and NH<sub>4</sub><sup>+</sup> removal from the standard MM medium on the total net flux through the system and the Glc demand for production of planktonic biomass in the presence of O<sub>2</sub>**

<i>In silico</i> N sources available	Reactions in solution (n)	Objective value (mmol/gDW/h)	Variation (%)	Glc consumed (mmol/gDW/h)	Variation (%)
All amino acids in MM medium	291	259	-	1.04	-
Ala, Arg, Glt	328	310	19.7	2.09	101
Ala	341	335	29.2	2.04	94.5
Arg	328	329	26.8	3.19	204
Glt	330	320	23.6	2.21	111
Mixture of amino acids non-suitable as N sources	-	∞	-	-	∞
NH <sub>4</sub> <sup>+</sup>	343	397	53.1	8.50	712

The variation on the objective value and the amount of Glc consumed was calculated as a percentage over the values obtained with the standard MM medium (291 mmol/gDW/h = 100% and 1.04 mmol Glc (/gDW/h) = 100%). Rows in orange correspond to conditions that showed growth experimentally. Rows in blue correspond to conditions where growth was not observed.

Although growth was observed in MM medium base and a mixture of Ala, Arg and Glt, these cultures presented an increased lag phase and a maximum  $A_{600}$  significantly lower at the 95% confidence level (P=0.003) than cultures grown in standard MM medium. When the corresponding *in silico* responses were compared, the solution associated with growth on the mix of amino acids presented an objective value (total net flux through the system) 19.7% higher than the solution obtained under standard conditions, and the Glc demand increased by 101%. For the remaining conditions tested, no growth was observed *in vitro*, while the *in silico* increase observed on these parameters was higher than on the amino acid mix, with the sole exception of growth on Ala, for which the Glc demand was lower.

## 5.4 Discussion

### 5.4.1 *In silico* utilisation of amino acids for biomass production

#### 5.4.1.1 Overall uptake and excretion of amino acids and their uptake to demand ratio for biomass production

The LP-based analysis results obtained for production of planktonic biomass while satisfying the GAM and NGAM ATP demand showed how several of the 19 amino acids available in the standard MM medium were taken up and catabolised for this purpose, with their consequent deamination leading to excretion of  $\text{NH}_4^+$ . The following 11 amino acids: Asn, Cys, His, Ile, Leu, Phe, Pro, Ser, Trp, Tyr, Val presented uptake to demand ratios of 1, indicating their direct incorporation to the biomass protein without being metabolised further. Full catabolic pathways are absent for some of them (Ile, Leu, Phe, Trp and Tyr), which helps explain their lack of use for anabolic purposes. These results partly coincide with the *in vitro* observations for amino acid utilisation in *S. aureus* described by Halsey *et al.* (Halsey *et al.* 2017) (Chapter 0, Section 1.6.2), which indicated that 6 of those 11 amino acids (Cys, Ile, Leu, Phe, Tyr and Val) were taken up and directly incorporated into biomass protein. Study of the fate of sulphur containing amino acids showed that, while Cys was not catabolised, a small amount of Met was utilised to obtain Asp via homo-Ser in a process that reduced NAD and NADP, which was advantageous, since these reduced cofactors could then be utilised in the ETC to obtain ATP (NADH) or in biosynthetic processes (NADPH). Amino acids with uptake to demand ratios of 0 were not imported from the medium, but instead produced by the system from other amino acids. This was the case of Gln and Gly: since Gln was not present in the medium it could only be incorporated to the biomass if synthesised from Glt. Gly, on the other hand, was supplied in the medium, but synthesising it as a by-product carried a lower associated total net flux than its uptake. Study of the reactions in the LP solution showed that Gly was obtained during the degradation of Thr in a reaction catalysed by a Thr-aldolase (EC 4.1.2.5). This process generated acetald, which could in turn be used for Ac-CoA and NADH production via the acetald-dh (EC 1.2.1.10). Finally, those amino acids with uptake to demand ratios  $>1$  were taken up by the system and catabolised to generate other compounds. These were, in decreasing order of utilisation: Glt, Ala, Thr, Asp, Arg, Lys and Met. The level at which the system catabolised Lys and Met was fairly low. The catabolism of Glt, Ala, Thr, Asp and Arg, has been described as important, not only for the synthesis of other amino acids, but also for central C metabolism processes and energy production: Glt catabolism is known to provide 2-KG to fuel the TCA cycle and gluconeogenesis while generating NADH for ATP synthesis via oxidative phosphorylation (Tynecka *et al.* 1999; Halsey *et al.* 2017). Arg is usually utilised as a precursor for Glt while Ala and Thr (together with Gly and Ser)

are catabolised to Pyr to be used for ATP synthesis in the Pta/AckA pathway, leading to Ac production in a process that also generates NADH (Halsey *et al.* 2017). Finally, Asp seems to be important for the production of oxaloacetate in order to replenish the TCA cycle (Halsey *et al.* 2017). Therefore, the high level of utilisation of Glt, Ala, Thr, Asp and Arg to support growth *in silico* accurately reproduced the main features of the biological behaviour described for staphylococci growing under similar conditions.

#### ***5.4.1.2 Contribution of individual amino acids to the total nitrogen and carbon uptake***

These data showed how Glt contributed to the total N imported for growth in a significantly higher proportion than any other amino acid, accounting for approximately 56% of the total N demand, followed by Ala, which accounted for approximately a 12%. While Arg and Thr each accounted for over a 6%, all remaining amino acids did so in percentages < 5% and in the following order (from higher to lower): Lys, Asp, Asn, Leu, Ile, His, Val, Ser, Phe, Tyr, Pro, Met, Trp and Cys (Section 5.3.1.2, Table 5-3). Taking these values into account, it is clear that Glt was the main amino donor in the anabolism of other amino acids.

The results obtained for the contribution of amino acids to the total C demand for growth were in line with those obtained for N, being relative to the number of C atoms present in each amino acid molecule. Glt and Ala again contributed to this demand in the higher proportions: Glt accounted for approximately the 60% and Ala for the 7% of the total C demand, followed by Thr (5%) and Asp (3%) etc. These data was also in line with the phenotypic utilisation of amino acids for growth described for staphylococci, where these are used to obtain energy and 2-KG, Pyr and oxaloacetate as key metabolic intermediaries in biomass synthesis (Tynecka *et al.* 1999; Halsey *et al.* 2017).

Comparison of model responses for growth in presence and absence of amino acids showed that when Glc was the only C source available, its metabolism led to production and excretion of Ac, CO<sub>2</sub> and Form. In this case, production of Ac and CO<sub>2</sub> indicated that the ATP demand for growth was being covered by Glc oxidation to Ac following the mechanism described in Chapter 3, Section 3.3.2.1,

Figure 3-3. Indeed, investigation of the LP solution showed how in this response ATP was generated together with Ac via Ac-CoA by the Ac-CoA ligase, while Form was generated as a by-product during the synthesis of co-factors (FAD). When Glc and all 19 amino acids in the MM medium were available, the C uptake increased by 55.9 %, and, in consequence, so did the excretion of metabolic by-products. In this case, the system produced CO<sub>2</sub> and Suc, suggesting Glt utilisation for ATP

production by the mechanism described in Chapter 3, Section 3.3.2.1, Figure 3-8. Again, investigation of the LP response confirmed this hypothesis, showing how Glt was catabolised to 2-KG, Suc-CoA and finally Suc, generating ATP via the Suc-CoA synthetase. These results suggested that ATP production is coupled with amino acid catabolism even in the presence of Glc when these are being utilised for growth support.

The total N and C demands for growth when  $\text{NH}_4^+$  was available as sole N source were much lower than in the presence of the 19 amino acids supplied with the standard MM medium, with these figures rising by 133% and 55.9 % respectively (Section 5.3.1.2, Table 5-4 and Table 5-6). In this case, all N was taken up as amino acids and 9.64 mmol/gDW/h were excreted as  $\text{NH}_4^+$  as a result of amino acid catabolism. These results showed that although higher amounts of N and C were consumed when biomass was produced from Glc and amino acids rather than from Glc and  $\text{NH}_4^+$ , the total net flux through the system decreased by 53.1% when amino acids were catabolised, and thus this was prioritised as an optimal solution (Section 5.3.2.5, Table 5-12). This implied following different metabolic routes, which finally led to an increase in the N and C demands and in the excretion of metabolic by-products. Furthermore, the total number of reactions involved in this response was also lower. Therefore, considering either total net flux or total number of reactions in the solution as proxies for enzymatic investment, these results indicate that amino acid uptake and catabolism for growth would be favoured in situations where this matches the real biological objective of the cell. A more detailed investigation of the reactions involved in these networks is out of the scope of this study.

#### **5.4.1.3 Growth on single amino acids as nitrogen sources**

From the 7 amino acids that cannot be utilised by the system as single N sources (Section 5.3.1.1, Table 5-7), 6 lack complete catabolic pathways (Ile, Leu, Lys, Phe, Trp and Tyr), explaining why they could not be used to support growth *in silico*. Since Glt was the main amino acid taken up from the medium and contributed in a higher proportion to the total N and C demand for biomass production (Section 5.3.1.2), it was expected that its use led to the solution with a lower objective value, while use of  $\text{NH}_4^+$  as sole N source had the highest associated objective value. This makes sense, since synthesising amino acids *de novo* involves the activity of a higher number of enzymes, being a more protein costly process than synthesising them from other amino acids.

#### **5.4.1.4 Nitrogen assimilation: glutamate biosynthesis from $\text{NH}_4^+$**

The responses obtained for Glt synthesis from  $\text{NH}_4^+$  seemed to match standard bacterial behaviour: the most common mechanisms of bacterial N assimilation involve the action of the Glt-dh (Glt

dehydrogenase) or the GS/GOGAT (Gln synthetase/Glt synthase) cycle. For example, *E. coli* has been shown to primarily assimilate  $\text{NH}_4^+$  via action of the Glt-dh, until the  $\text{NH}_4^+$  concentration available falls below 1mmol, causing the affinity of the enzyme for this substrate to decrease and the GS/GOGAT system to take over (Amon *et al.* 2010). While the optimal LP-based analysis response obtained for Glt synthesis has not been specifically described in staphylococci, it still involved the action of a Glt-dh and the GS. Minimisation of the total net flux through the system was achieved by also allowing flux through other reactions (e.g. aminotransferases). The solution identified upon blocking flux through the Glt-dh reactions involved the GS/GOGA cycle, which accurately reproduced this bacterial strategy for Glt-synthesis. These results also showed how even though Glt synthesis would preferentially involve a cycle of production and consumption of Glt and Gln, catabolising Gln is not essential for this process. However, the action of either a Glt-dh or a GOGAT enzyme (NAD- or NADP-dependent) is always required, thus Glt synthesis always involves utilising 2-KG as its precursor.

## **5.4.2 Experimental work for model validation on amino acid utilisation for biomass production**

### ***5.4.2.1 Assessment of growth in the absence of proline***

These experimental results finally proved that RP62A is capable of growing without Pro as long as sufficient amounts of ornithine-precursor amino acids (Arg, Gln and Glt) are available in the media. The ability of these organism to synthesise Pro was correctly described by the GSM in the first place, and here we speculate that the lack of growth observed on MM<sup>-</sup> medium without Pro was due to a limited availability of ornithine precursors, which impaired growth significantly and seemed to trigger a stress response that induced biofilm formation: although growth in MM<sup>-</sup> medium lacking Pro was not detected even after 7 days of incubation, cells still survived on these samples, and colonies could be recovered by the end of the experiment after their incubation on BHI agar. A small amount of bacterial biomass could be detected by eye on the bottom and the walls of the tubes, which could be compatible with the production of EPS and the formation of a biofilm, indicating that Pro deprivation under these conditions triggered phenotypic changes in the cells. A similar behaviour has been reported in staphylococci in response to amino acid starvation or other adverse conditions in what is known as the stringent response (Lister *et al.* 2014; Somerville 2016). This could explain the maximum  $A_{600}$  value of  $0.11 \pm 00$  SD observed in these cultures during the experiments described here and in Chapter 4. The possible effect of catabolite C repression on these cultures was also considered, since Glc is present on samples formulated from MM medium base. However, both the medium containing a mixture of Ala, Arg and Glt as sole N source and the MM<sup>-</sup> medium without

Pro contained equal amounts of Glc, therefore, the regulatory effects derived from its presence should be similar in both cases. Thus it seems unlikely for the growth observed after 5 days of incubation on samples formulated with the mixture of Ala, Arg and Glt to be due to the de-repression of amino acid catabolism upon Glc consumption since a) no growth was observed before that time point, hence there is no reason to believe that the Glc concentration would have fallen by day 5, and b) no growth was observed on samples containing MM<sup>-</sup> medium without Pro even after 14 days of incubation. Hence the difference in concentration of ornithine-precursor amino acids between these two media remains the most plausible explanation for the differences in growth observed *in vitro* and explains the discrepancy detected between the positive *in silico* result for growth in the absence of Pro and the *in vitro* lack of growth in MM<sup>-</sup> medium lacking this amino acid.

#### ***5.4.2.2 Assessment of growth on single amino acids (alanine, arginine and glutamate) and on a mixture of them***

Growth in the presence of these three amino acids combined was possible but cultures presented a lower maximum  $A_{600}$  value than those on standard MM medium and an extended lag phase. This indicated that even though growth was supported under these conditions, cells were dividing at a lower rate. Here we hypothesise that this decrease in growth could be a result of cells having to adapt to an amino acid-deficient environment, where plenty of biosynthetic enzymes would have to be either synthesised *de novo* or re-activated in order to allow production of every other amino acid lacking in the medium. As explained before, staphylococci seem to utilise these amino acids in the following manner (Halsey *et al.* 2017): while Ala can be catabolised for ATP synthesis producing Pyr and subsequently Ac, Glt enables functioning of the TCA cycle generating metabolic intermediates prior conversion to 2-KG, generating ATP and NADH in the process. Finally, Arg can be converted to citrulline, allowing production of Pro via ornithine, providing that its catabolism is not repressed by the presence of Glc in the media. Therefore, it is possible that, when these three amino acids are provided in sufficient amounts, cells are able to obtain enough biosynthetic precursors and energy to grow. Here we speculate that the lack of growth observed *in vitro* when these amino acids were provided individually, even after an extended incubation period of 14 days, could be due to cells not being able to cover the increased enzymatic cost and/or Glc demand associated to responses for growth under these conditions. Furthermore, since the intracellular space is limited and constitutes a vital resource that must be allocated strategically, there might not be enough free space to allow the simultaneous synthesis and functioning of all the enzymes required for these responses. A similar concept has been used to explain certain bacterial metabolic responses such as bacterial ‘overflow’ metabolism, which was proposed to be induced by protein membrane overcrowding (Kizil 2010). Further investigation of these responses is out of the scope of this study.

### **5.4.2.3 Assessment of growth on a mixture of amino acids defined as non-suitable nitrogen sources for RP62A and on $\text{NH}_4^+$ as sole nitrogen source**

As expected, the inability of the organism to grow solely in the presence of the 7 amino acids defined by LP-based analysis as non-suitable N sources was also confirmed *in vitro*. However, the lack of growth on samples containing  $\text{NH}_4^+$  as sole N source contrasted with the system's ability to reproduce growth *in silico* under these conditions. This could again be explained as either a consequence of: i) the severe increase on the total net flux through the system (+ 53.1%) and/or the increase in the Glc demand (+712%) associated with this response in comparison to growth on standard MM medium, and/or, ii) as a result of the effect of regulation on gene expression for enzymes involved in the biosynthesis of amino acids (Gladstone 1937; Knight 1937; Heinemann *et al.* 2005; Lee *et al.* 2009; Bosi *et al.* 2016) (Chapter 0, Section 1.5).

### **5.4.2.4 The effect of Glc on amino acid metabolism**

While catabolite C repression has been described in staphylococci at similar Glc concentrations as the ones in the media used for this study (13.9 mmol (Halsey *et al.* 2017), 25.8 mmol (Sivakanesan *et al.* 1980), and 51.6 mmol (Tynecka *et al.* 1999) vs 11.1 mmol, 22.2 mmol and 111 mmol), its involvement in the *in vitro* responses observed here is unclear. Furthermore, we identified certain level of controversy between studies reporting the effects of C catabolite repression on staphylococcal amino acid metabolism, with some authors describing how the presence of Glc inhibited the metabolism of Arg and Pro but not that of Glt or Ser (Halsey *et al.* 2017), while others reported Glc to stop Glt catabolism (Tynecka *et al.* 1999) or Ser utilisation (Sivakanesan *et al.* 1980). On top of this, the biological objectives behind this phenomenon are partly obscure: while it is reasonable to assume that low amino acid concentrations would lead to a decrease in their metabolism and their direct incorporation into the biomass, thus favouring Glc consumption and utilisation of  $\text{NH}_4^+$  for anabolic processes and energy production, the reasons behind amino acid catabolic repression when amino acids are also available in excess are not so intuitive. Possible theories for this suggest that, since the intracellular volume is limited and Glc storage polymers are big in size, Glc consumption is maintained inside the cell as a way to control its intracellular concentration, effectively managing allocation of the intracellular space. Since enzyme kinetic parameters and complete transcriptomic datasets are hard to integrate in GSMs, these do not take into account the possible metabolic effects of catabolite C repression, which is one of the current limitations of genome-scale metabolic modelling. The lack of accurate information available regarding transporters for RP62A led to the inclusion of transporting reactions that do not take into account the symport of

compounds or the ATP demand associated with some of these processes. Thus since transport of certain amino acids could involve ATP consumption and import of Glc by the PTS system can effectively save some of the ATP needed for its phosphorylation during glycolysis, results derived from the analysis of the GSM could be overestimating the utilisation of amino acids over Glc when both are present in the *in silico* media. However, these are still useful to understand what will occur in media with a low Glc concentration or upon its depletion during the post-exponential growth phase.

## 5.5 Conclusion

The analyses described in this chapter have provided further insight into the system's N and amino acid metabolism. The conclusions that can be derived from the validation work presented here are summarized in Table 5-13 and are the following:

- i. As correctly described by the GSM in the first place, growth without Pro is possible when the medium contains sufficient amounts of ornithine precursors, such as Arg or Glt. The absence of growth detected during previous experimental work (Chapter 4) was presumably due to low concentrations of these amino acids in the medium and not to an error in the model. Feeding ornithine to cultures in MM<sup>-</sup> without Pro could have been a way to test the ability of RP62A to produce this amino acid from its direct precursor.
- ii. Growth in the presence of the three amino acids which contribute in the highest proportion to the total N uptake for biomass production (Ala, Arg and Glt) is possible, both *in silico* and *in vitro*. However, growth in the presence of these three amino acids supplemented independently or on NH<sub>4</sub><sup>+</sup> as sole N source was solely observed *in silico*. This discrepancy could be explained as a result of the increased Glc demand and total net flux associated with growth upon these conditions, and/or, as a result of the effect of regulation on genes involved in amino acid metabolism. If the assumption made throughout this work that considers total net flux through the system as a proxy for protein investment is accurate, these results would link growth under these conditions with a higher enzymatic cost and, therefore, an increased demand for cellular space that the biological organism might or might not be able to meet.
- iii. Absence of growth in a combination of the 7 amino acids defined by LP-based analysis as non-suitable N sources matches the behaviour of the organism *in vitro*.
- iv. Production of the energy needed for growth could be tightly coupled with amino acids catabolism even when Glc is present in the media.

**Table 5-13 Comparison of results derived from LP-based analysis of the model and the experimental work performed during this study**

<b>Amino acids auxotrophies and utilization of N sources</b>	<b>Experimental results</b>	<b>Model LP-based analysis results</b>	<b>Agreement between <i>in silico</i> and <i>in vitro</i> results</b>	<b>Proposed explanation for disagreements</b>
<b>Essential amino acids</b>	None	None	Yes	NA
<b>Growth in a mixture of Ala, Arg and Glt</b>	Growth observed	Growth allowed	Yes	NA
<b>Growth in Ala as sole N source</b>	No growth	Growth allowed	No	Excessive Glc demand, protein cost and/or regulation of gene expression
<b>Growth in Arg as sole N source</b>				
<b>Growth in Glt as sole N source</b>				
<b>Growth in amino acids that cannot be utilized as single N sources for growth</b>	No growth detected on these amino acids combined	No growth	Yes	NA
<b>Growth with NH<sub>4</sub><sup>+</sup> as sole N source</b>	No growth	Growth allowed	No	Excessive Glc demand, protein cost and/or regulation of gene expression

NA = Non-applicable

This work exemplifies how the results derived from LP-based analysis of the model can be used to formulate hypothesis regarding certain characteristics of the organism (e.g. RP62A being capable of biosynthesizing Pro) and can successfully be applied to the design of experiments that test them *in vitro*.

Overall, the results shown in Chapters 3 to 5 present a good level of coincidence between the *in vitro* data and the behavior of the model, with the disagreements encountered explained in a reasonable manner. This, together with the high level of manual curation applied, justifies using this GSM as a guiding tool for future experimental work and, to a certain extent, the assignment of biological significance to predictions derived from its analysis.

## 6 Applying model analysis to the study of the metabolism of RP62A cells growing on prosthetic joints

### 6.1 Introduction

Staphylococcal cells colonising prosthetic joints do so by forming biofilms on the surface of the implants and/or biological structures present in the intra-articular space (Chapter 0, Section 1.8). The metabolic shifts that allow cells to switch from production of planktonic biomass to synthesis of biofilm components, and to adapt to the O<sub>2</sub> gradient encountered across biofilm structures (Wimpenny *et al.* 1983; Wimpenny *et al.* 2000) are still largely unknown. This chapter applies LP-based analysis to the study of these strategies and defines potential metabolic routes involved in biofilm formation in *S. epidermidis* RP62A:

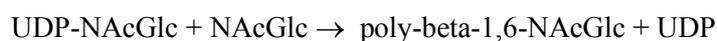
Since it is not practical to tackle production of several biofilm types in a single study (Chapter 0, Section 1.7), this work focuses on how PIA, the main biofilm matrix component of *ica*-positive strains, is biosynthesised by RP62A in the intraarticular space (Chapter 0, Sections 1.7.3 and 1.7.4). This way, LP was used to analyse the model for production of energy (ATP), planktonic biomass and PIA in synovial fluid, considering an array of O<sub>2</sub> concentrations and in the presence and absence of NO<sub>3</sub><sup>-</sup>, thus replicating the conditions that cells encounter when growing as part of biofilms on prosthetic joints: initially, the system was analysed to obtain single solutions for the most extreme conditions studied, thus in full presence or absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>. Once these single solutions were investigated, the model was ‘scanned’ (Chapter 2, Section 2.2.4) for production of ATP, planktonic biomass and PIA under a range of O<sub>2</sub> concentrations, thus mimicking the O<sub>2</sub> gradient found across biofilm layers. The reactions included in the scan solutions were extracted and used to generate submodels. Finally, since the net behaviour of a system can be expressed as a linear combination of elementary modes, these were computed on the reduced networks, thus defining all feasible minimal routes through the subsystems derived from the scan responses and allowing a more detailed investigation of these data.

Biofilms are produced as part of a response to stress-inducing environmental factors (Sadykov *et al.* 2008; Somerville 2016). Therefore, it is reasonable to assume that reducing the protein costs for cellular growth and survival is a plausible biological objective for bacteria in this phenotypic state. Thus the objective function chosen for the analysis presented in this chapter is minimisation of total net flux through the system.

### 6.1.1 Synthesis of PIA in the RP62A GSM

The PGDB for RP62A used for model construction did not contain any of the reactions associated with the genes in the *ica* operon (Chapter 0, Section 1.7.4). However, these genes were identified during the annotation of the RP62A genome and were automatically assigned metabolic functions by protein homology with the *E. coli* genes of the *pgaABCD* locus (Wang *et al.* 2004), which promotes the synthesis of a polysaccharide adhesin similar to PIA, and required for biofilm formation. The functions assigned to these genes were: SERP\_RS11285 = poly-beta-1,6-NAcGlc synthase (IcaAD); SERP\_RS11295 = poly-beta-1,6-NAcGlc deacetylase (IcaB); SERP\_RS11300 = poly-beta-1,6-NAcGlc export plasma membrane protein (IcaC); SERP\_RS11290 = poly-beta-1,6-NAcGlc synthesis protein (IcaD); and SERP\_RS11280 = biofilm operon *icaADBC* transcriptional regulator (IcaR). Thus a set of reactions for PIA synthesis were manually included in the model:

- i. According to MetaCyc, IcaA presents a 35% amino acid identity and 57% similarity with the PgaC protein of *E.coli*. The *pgaC* gene encodes for a poly-beta-1,6-NAcGlc synthase (EC 2.4.1.-) that catalyses the polymerization of single monomer units of UDP-NAcGlc to produce the linear polymer poly-beta-1,6-NAcGlc through the reaction described below:



In order to avoid introducing material inconsistencies due to the inclusion of metabolites with undefined empirical formulae, as is the case of NAcGlc-based polymers, the reactions involved in the direct synthesis of PIA were described following the logic explained in Chapter 2, Section 2.4.4.3. Thus polymers were assigned the formulae of their monomeric units and the stoichiometries of the reactions were adjusted accordingly. This way, an initial PIA synthesising reaction was defined as follows:



where PIA1 represents poly-beta-1,6-NAcGlc, with the same empirical formulae as NAcGlc. A molecule of H<sub>2</sub>O and a proton had to be included on this reaction in order to assure its atomic balance.

- ii. The *icaB* gene is homologous to the *pgaB* gene in *E. coli*, which encodes for a poly-beta-1,6-NAcGlc deacetylase. According to BRENDA, the EC 3.1.1.59 is associated with these genes, and corresponds to the catalysis of the following reaction:



thus a deacetylating reaction for PIA was defined as follows, and included in the model:



where PIA2 represents a partially deacetylated NAcGlc polymer with the same empirical formulae as one unit of glucosamine.

- iii. Finally, the homology of the *icaC* gene with the *E. coli pgaD* and *pgaC* genes involved in the transport of poly-N-NAcGlc to the periplasmic space was used as a base to introduce exporters for PIA1 and PIA2. Since this occurs without an associated ATP cost in *E. coli*, the PIA1 and PIA2 exporters were also described as non-energy consuming reactions.
- iv. Finally, the *pgaA* gene in *E. coli* codifies for an outer membrane porin, with no functional homology in *S. epidermidis* since this is a Gram-positive organism. Therefore, no further reactions were included in the system.

### 6.1.2 Composition of the *in silico* synovial fluid

According to the composition of the organic synovial fluid described in Chapter 0, Section 1.7.2.1, an *in silico* version of it could be defined as a medium containing the same components as the *in silico* standard MM medium (Chapter 4, Section 4.2.3.5) with the following additions: aminobutyric acid, citrulline, hyaluronate, ornithine, taurine, urea, ureate and  $\text{NO}_3^-$ .

Currently, there is no evidence of *S. epidermidis* RP62A being able to generate the enzyme hyaluronidase, a virulence factor of *S. aureus* which can break hyaluronic acid into glucuronic acid and NAcGlc. According to MetaCyc vs 23.0 and BRENDA vs 2019, the two main bacterial enzymes catalysing reactions where NAcGlc is consumed are EC 2.7.1.221 (N-acetylmuramate-1-kinase) and EC 2.7.1.162 (N-acetylhexosamine-1-kinase), both of which phosphorylate NAcGlc, producing NAcGlc-1P while consuming ATP. These enzymes have only been described in *Pseudomonas putida* and *Bifidobacterium longum* (respectively), and seem to be absent in *S. epidermidis*. In addition, no evidence of *S. epidermidis* RP62A being able to metabolise taurine or ureate was found. Taking this into account, only five of the eight compounds initially added to the standard MM medium in order to describe an *in silico* version of synovial fluid could potentially be utilised by the system (aminobutyric acid, citrulline, ornithine, urea and  $\text{NO}_3^-$ ). Therefore, these were the only compounds finally included in the *in silico* synovial fluid considered during the analyses described in this chapter.

## 6.2 Methods

### 6.2.1 Model analysis for production of ATP, planktonic biomass and PIA in joints

#### 6.2.1.1 Metabolic responses for ATP and biomass production in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>

LP-based analysis was utilised to obtain model responses for three different processes: i) production of 45 mmol/gDW/h of ATP to cover the NGAM and GAM cell's energy demand (Chapter 3, Section 3.3.4), ii) synthesis of PIA (80% PIA 1 and 20% PIA2), and iii) synthesis of planktonic biomass. The LP formulations used were defined as follows (Equation 6-1):

$$\begin{array}{l} \text{minimize:} \quad \sum_{i=1}^m |v_i| \\ \\ \text{subject to} \quad \left\{ \begin{array}{l} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ \text{i)} \quad \mathbf{v}_{\text{ATPase}} = \text{ATPase} = 45 \\ \text{ii) and iii)} \quad \mathbf{v}_j \geq t_j ; k \leq j \leq m \end{array} \right. \end{array}$$

where the notations and constraints are the same as in Equation 6-1 (Chapter 3, Section 3.2.1) with the following modifications: the flux through the ATPase reaction and reactions of  $N$  corresponding to exporters for biomass components (from  $k$  to  $m$ ) were specified to allow production of either i) 45 mmol/gDW/h of ATP, ii) 1 gram of PIA or iii) 1 gram of planktonic biomass. The export of compounds included in the composition of either planktonic biomass or PIA was enabled to vary in order to reproduce biological states closer to those found *in vivo*, where organisms may modify their biomass composition in response to changes in the environmental conditions, following the method previously described by others (Villanova *et al.* 2017). Thus the upper flux bounds of the constraints for these exporters were relaxed. Further constraints were added if the analysis so required (e.g. blocking flux through the O<sub>2</sub> and/or the NO<sub>3</sub><sup>-</sup> importer). Finally, the percentage in which the consumption of each medium component contributed towards the total amount of C and N taken up in the process was calculated.

#### 6.2.1.2 Simulated variation in the O<sub>2</sub> concentration during ATP and biomass production

With a focus on studying individually model coordinated responses for either production of energy,

PIA or planktonic biomass in response to changes in the O<sub>2</sub> level, the LP-based analysis technique was applied as follows: the solution space was repeatedly scanned over increasing values for a given constraint (here O<sub>2</sub> availability), as previously described by others (Poolman *et al.* 2009; Villanova *et al.* 2017). This method identifies those reactions which respond in a coordinated fashion to a changing constraint. The LP formulation utilised was defined in a similar manner as in Equation 6-1, with an additional constraint (Equation 6-2):

$$\begin{aligned} &\text{minimize:} && \sum_{i=1}^m |\mathbf{v}| \\ &\text{subject to} && \left\{ \begin{array}{l} \mathbf{N}_{n,m} \cdot \mathbf{v} = \mathbf{0} \\ \mathbf{v}_{O_2\_importer} \leq O_{2limit} \\ \text{i) } \quad \mathbf{v}_{ATPase} = ATPase = 45 \\ \text{ii) and iii) } \quad \mathbf{v}_j \geq t_j ; k \leq j \leq m \end{array} \right. \end{aligned}$$

here,  $\mathbf{v}_{O_2\_importer} = \leq O_{2limit}$  defines the upper flux limit for the O<sub>2</sub> importer. This parameter increased in a linear fashion throughout the scan, ranging from zero to 12 mmol/gDW/h, value above which all reactions responding to the change in O<sub>2</sub> concentration exhibited a constant flux. Again, further constraints were added if needed (e.g. block of NO<sub>3</sub><sup>-</sup> import or blocking flux through reactions involved in multiple optimal solutions).

### 6.2.1.3 Generation of submodels and computation of elementary modes

All reactions included in the dataset of solutions computed in each O<sub>2</sub> scan were extracted and used to generate submodels. Therefore, each submodel included all reactions involved in the production of either ATP, PIA or planktonic biomass under the range of O<sub>2</sub> concentrations tested. While some carried flux at a fixed rate, others varied their flux significantly throughout the scan. A change in flux indicates that these reactions are actively responding to the variation in the O<sub>2</sub> concentration, which highlights their importance in the metabolic adaptation from aerobic to anaerobic conditions. Finally, the reduced number of reactions contained in each subnetwork allowed definition of all the stoichiometrically feasible minimal routes through these systems by computing their elementary modes following techniques previously described by others (Poolman *et al.* 2009; Villanova *et al.* 2017). The large number of reactions in the submodels generated from results obtained with the O<sub>2</sub> scans for production of planktonic biomass did not allow the computation of elementary modes through these networks.

Relevant code to the work performed in this chapter can be found in Appendix A, Section 9.1.3.3.

## **6.3 Results**

### **6.3.1 Production of ATP to cover the GAM and NGAM energy demand**

The results obtained applying the methods described in Section 6.2 to the study of ATP production in joints are described below.

#### ***6.3.1.1 Metabolic responses for ATP production in synovial fluid in the presence and absence of $O_2$ and $NO_3^-$***

Single solutions for ATP production under conditions reproduced by the constraints applied in the extreme points of the  $O_2$  scans, thus under either total availability or absence of  $O_2$  and  $NO_3^-$ , were computed and the results obtained are presented in this section (Figure 6-1 to Figure 6-4). Greyed out reactions in these diagrams did not carry flux.

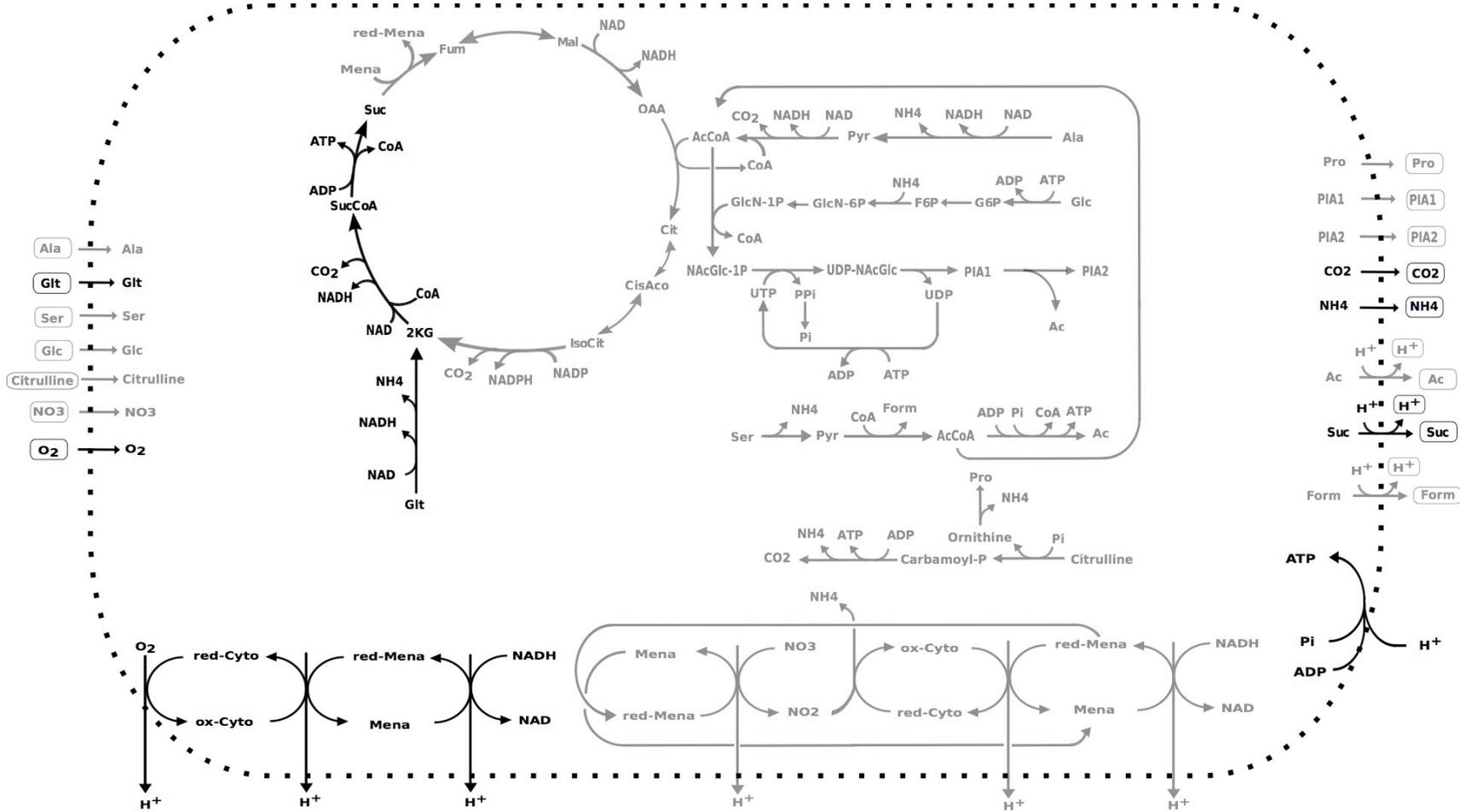


Figure 6-1 ATP production in synovial fluid in the presence of O<sub>2</sub> when total flux through the system was minimised.

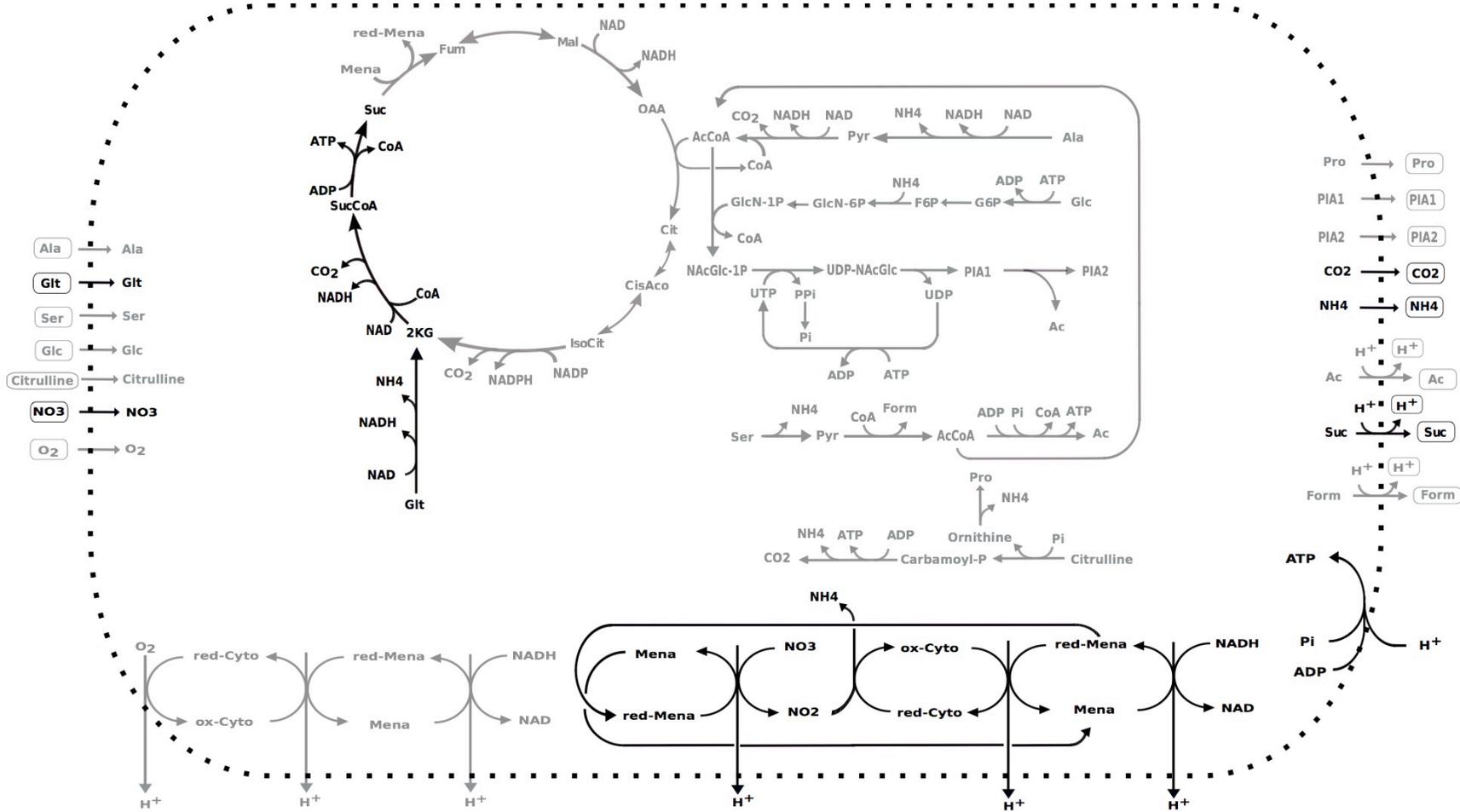


Figure 6-2 ATP production in synovial fluid in the presence of NO<sub>3</sub><sup>-</sup> when total flux through the system was minimised.

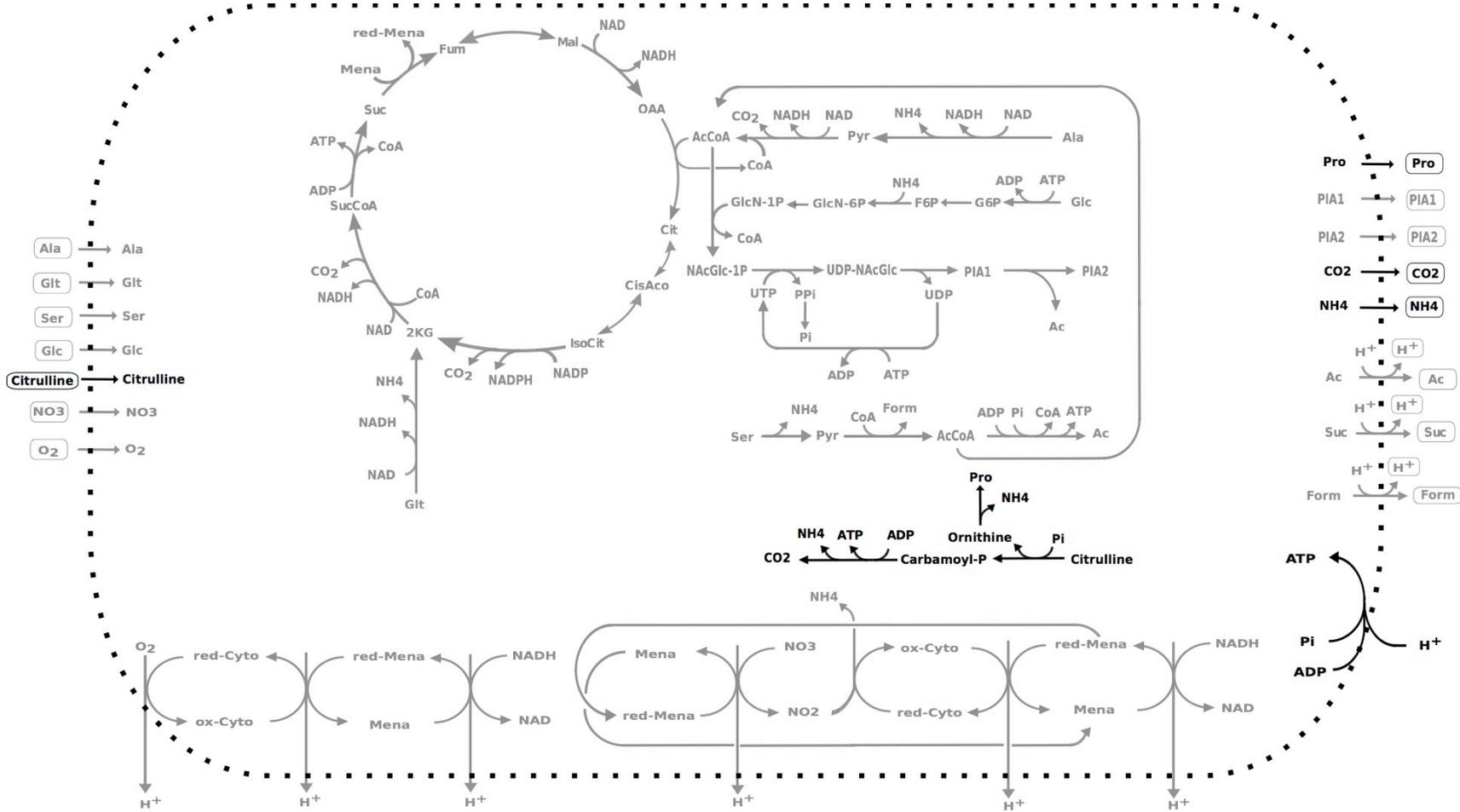


Figure 6-3 ATP production in synovial fluid in the absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> when total flux through the system was minimised.

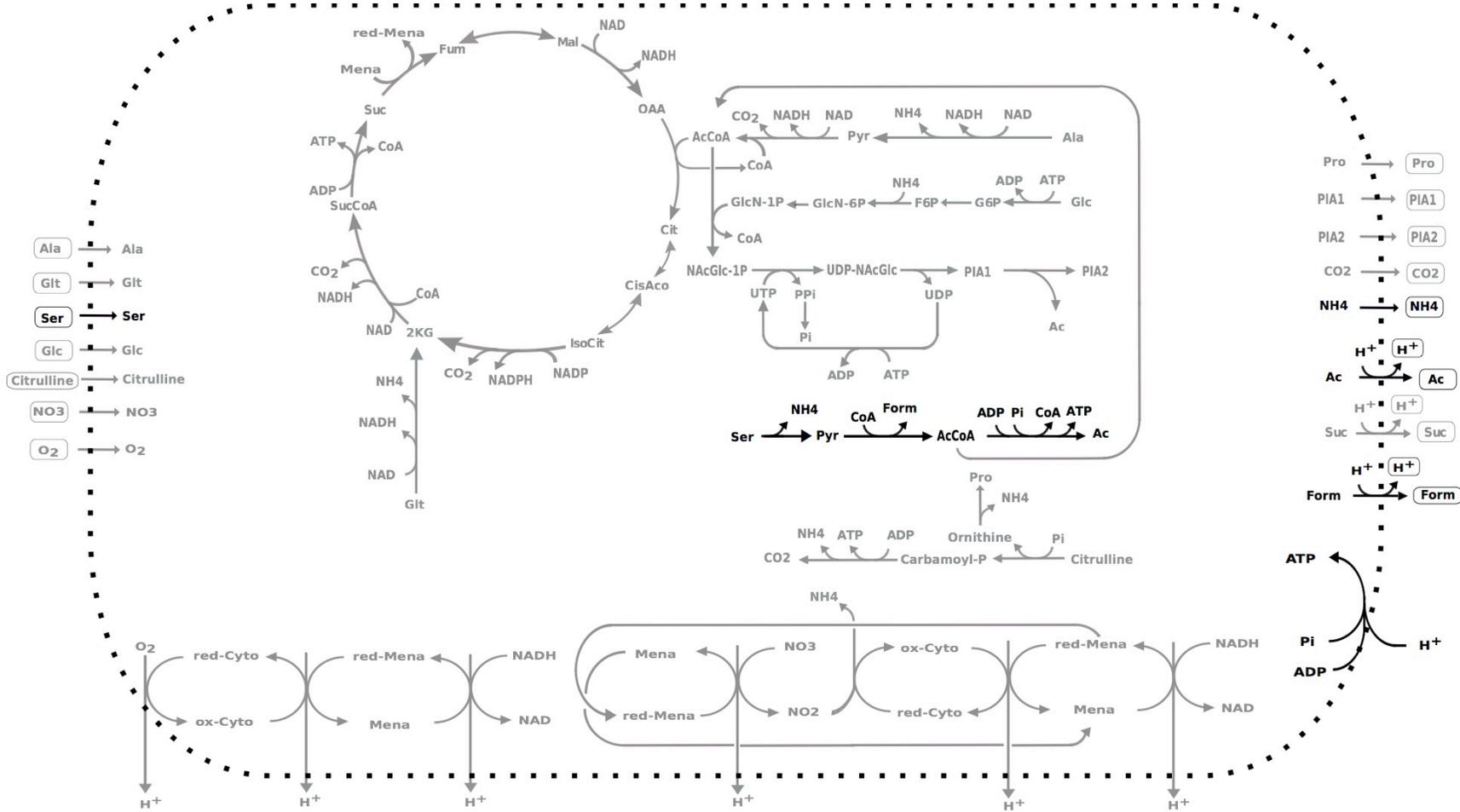


Figure 6-4 ATP production in synovial fluid in the absence of electron acceptors under depletion of citrulline when total flux through the system was minimised.

During production of ATP in synovial fluid in the presence of O<sub>2</sub> when total flux through the system was minimised (Figure 6-1) 2-KG was obtained from Glt via action of a NAD-dependent Glt-dh. 2-KG was then oxidised to Suc, generating ATP and NADH. This NADH entered the ETC during aerobic respiration, where NAD was regenerated. 1.1 molecules of ATP were produced per C (5.5 molecules per Glt) in this response. The protons excreted with Suc contributed to the PMF that pumps the ATP synthase.

When O<sub>2</sub> was absent but NO<sub>3</sub><sup>-</sup> was available, a similar response was observed (Figure 6-2), however, in this case NO<sub>3</sub><sup>-</sup> was utilised as the ultimate electron acceptor of the ETC. This solution generated 0.95 molecules of ATP per C (4.75 molecules of ATP per Glt), being slightly less efficient than following aerobic respiration. This response matches experimental observations where the menaquinone-dependent NO<sub>3</sub><sup>-</sup>-reducing mechanism of the ETC is favoured *in vitro* over its cytochrome-NO<sub>3</sub><sup>-</sup> reducing alternative (Uribe-Alvarez *et al.* 2016), as was previously described in Chapter 3, Section 3.3.2.1 during the study of ATP production from Glc.

When neither O<sub>2</sub> nor NO<sub>3</sub><sup>-</sup> were available (Figure 6-3), citrulline was taken up and metabolised to carbamoyl-P and ornithine by the ornithine carbamoyltransferase (EC 2.1.3.3), with these products being respectively degraded to CO<sub>2</sub>, NH<sub>4</sub><sup>+</sup> and Pro, and ATP being generated in the process. These compounds were then excreted from the system. It is important to emphasise that citrulline-ornithine antiporters have been described in eukaryotic cells (Bradford *et al.* 1980), but not in staphylococci. Therefore, the deamination of ornithine to Pro and its subsequent excretion may not be involved in this process. The protons excreted with NH<sub>4</sub><sup>+</sup> contributed to the PMF that pumps the ATP synthase.

In the absence of electron acceptors and citrulline (Figure 6-4), Ser was deaminated generating Pyr, which was then transformed to AcCoA by the Pyr Form-lyase (EC 2.3.1.54). AcCoA was subsequently utilised to generate ATP, producing Ac by the Ac-CoA ligase (EC 6.2.1.1) working in the reversible manner in a similar way as described before in solutions for ATP production from Glc (Chapter 3, Section 3.3.2.1, Figures 3-3, 3-4, 3-6 and 3-7). The protons excreted with these by-products contributed to the PMF that pumps the ATP synthase.

The main parameters defining these responses are summarised in Table 6-1 and Table 6-2:

**Table 6-1 Characterisation of the model behaviour for production of ATP in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.**

Electron acceptor available	Reactions carrying flux	Objective value	O <sub>2</sub> uptake	NO <sub>3</sub> <sup>-</sup> uptake	Total C uptake	Total N uptake	P/O ratio	ATP/C ratio
O <sub>2</sub>	13	188	8.18	0.00	40.9	8.18	2.75 2.25	1.10
NO <sub>3</sub> <sup>-</sup>	15	204	0.00	4.74	47.4	14.2	-	0.95
None	10	330	0.00	0.00	180	90.0	-	0.25

Units of the objective value and uptake rates = mmol/gDW/h. Conditions set: presence of O<sub>2</sub>, anaerobiosis and presence and absence of NO<sub>3</sub><sup>-</sup>. When both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> are available, O<sub>2</sub> is the only electron acceptor utilised, thus this condition was omitted from the table. For those conditions showing two P/O ratio values, the first value corresponds to the oxidative plus substrate level phosphorylation and the second one to the oxidative phosphorylation only.

**Table 6-2 Main N and C sources utilised and by-products excreted during production of ATP in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.**

Electron acceptor available	Compounds utilized	Contribution to total N uptake (%)	Contribution to total C uptake (%)	By-products excreted
O <sub>2</sub>	Glt	100	100	CO <sub>2</sub> , Suc
NO <sub>3</sub> <sup>-</sup>	Glt	66.7	100	CO <sub>2</sub> , Suc
	NO <sub>3</sub> <sup>-</sup>	33.3	-	
None	Citrulline	100	100	CO <sub>2</sub> , Pro

As expected, ATP production was achieved with a lower total net flux through the system by utilising O<sub>2</sub> as the final electron acceptor. When either of the electron acceptors were available, Glt was the preferred substrate utilised for ATP production, while citrulline took over when O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> were absent. The response obtained in the presence of O<sub>2</sub> is identical to that found when studying the synthesis of ATP from Glt as sole C source considering minimising total net flux through the system as the objective of the analysis (Chapter 3, Section 3.3.2.2). For comparison purposes, the reactions involved in this solution are represented in Figure 6-1.

The ability of the system to generate ATP in a far more restricted environment, thus in the absence of electron acceptors and upon depletion of urea and all free amino acids (including aminobutyric acid, citrulline and ornithine) was already tested in Chapter 3, Section 3.3.2.1 (Figure 3-6). For comparison with other results presented in this chapter the main features of this response are summarised in Table 6-3. This showed that although feasible, synthesising ATP from Glc instead of Glt increased the total net flux through the system by 12.4%.

**Table 6-3 Characterisation of the model behaviour for production of ATP in the absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> with Glc and NH<sub>4</sub><sup>+</sup> as sole available C and N sources when total net flux through the network was minimised.**

Compounds generated	Reactions carrying flux	Objective value	Increase in objective value (%)	Total C uptake	Total N uptake	ATP/C ratio
ATP	23	371	12.4	67.5	0.00	0.667

Units: mmol/gDW/h. The increase in the objective value refers to the value initially obtained in standard synovial fluid.

### ***6.3.1.2 Simulated variation in the O<sub>2</sub> concentration during ATP production: responsive reactions, subnetworks and elementary modes***

Once single solutions for ATP production under the most extreme conditions encountered by cells in joints were investigated, the system was analysed for production of ATP applying constraints that mimicked the O<sub>2</sub> gradient found across the layers of a biofilm. Following the method described in Section 6.2.1.2, the model was ‘scanned’ for production of ATP simulating a range of O<sub>2</sub> concentrations, thus allowing the identification of coordinated responses for production of energy in response to changes in the O<sub>2</sub> level. The results obtained are presented in this section. The main parameters defining these responses are summarised in Table 6-4:

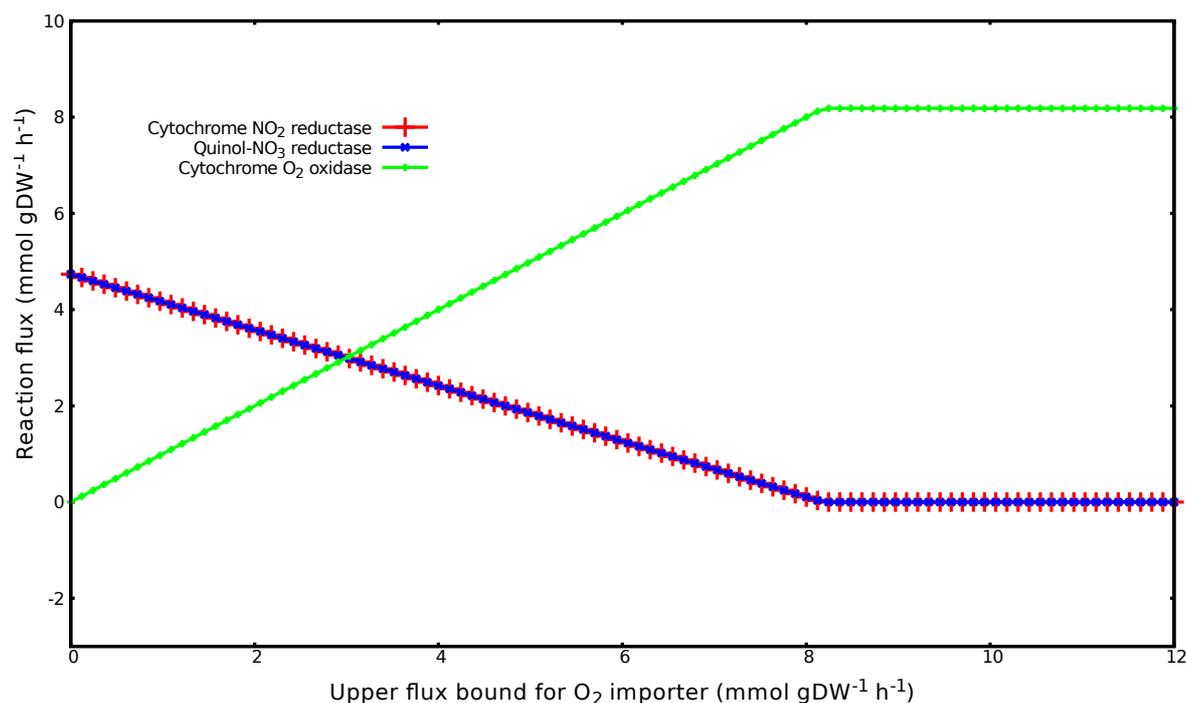
**Table 6-4 Summary of data obtained with O<sub>2</sub> scans for ATP production in synovial fluid in the presence and absence of NO<sub>3</sub><sup>-</sup> when total net flux through the network was minimised.**

NO <sub>3</sub> <sup>-</sup> available	Reactions active through the scans (n)	Changers (n)	Transporters in changers	EMs in submodels (n)
Yes	17	16	Glt, O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup> , H <sub>2</sub> O, NH <sub>4</sub> <sup>+</sup> , CO <sub>2</sub> , Suc	2
No	19	18	Glt, citrulline, O <sub>2</sub> , H <sub>2</sub> O, NH <sub>4</sub> <sup>+</sup> , CO <sub>2</sub> , Suc, Pro	2

Changers = reactions with a total flux variation above a threshold of 0.01 mmol (/gDW/h)

The only active reaction throughout both scans presenting fixed flux, hence not responding to changes in the upper flux bound for the O<sub>2</sub> importer, was the ATPase reaction.

Datasets obtained across model scans can be plotted, thus allowing for visualisation of the relative change in flux of the different reactions. For example, in order to investigate how reactions involved in aerobic and anaerobic respiration responded to changes in the O<sub>2</sub> availability when NO<sub>3</sub><sup>-</sup> was present in the *in silico* medium, their fluxes across the O<sub>2</sub> scan were plotted in Figure 6-5:



**Figure 6-5 Change in flux for responsive reactions involved in aerobic and anaerobic respiration during ATP production in synovial fluid over variation in the availability of O<sub>2</sub> when NO<sub>3</sub><sup>-</sup> was available and total net flux through the system was minimised.**

The values corresponding to the upper flux bound for the O<sub>2</sub> importer are plotted in the X axis while reaction fluxes are plotted in the Y axis. Units: mmol/gDW/h.

In the presence of NO<sub>3</sub><sup>-</sup>, when O<sub>2</sub> was completely absent, the O<sub>2</sub> importer and the cytochrome-c O<sub>2</sub> oxidase carried no flux, while flux through the NO<sub>3</sub><sup>-</sup> importer, the cytochrome-c-NO<sub>2</sub><sup>-</sup> reductase and the quinol-NO<sub>3</sub><sup>-</sup> reductase was maximum (approximately 5.00 mmol/gDW/h). Both sets of reactions changed flux in a mirror-like fashion, reaching equal fluxes of approximately 3.00 mmol/gDW/h, when the upper flux bound for the O<sub>2</sub> importer was set at 3.00 mmol/gDW/h. Finally, when this upper flux bound reached a value of 8.00 mmol/gDW/h, NO<sub>3</sub><sup>-</sup> was no longer utilised and fluxes through the O<sub>2</sub> importer and the cytochrome-c O<sub>2</sub> oxidase reached a plateau, staying at that level. This indicated that the objective value (total net flux through the network) of the solution obtained when the O<sub>2</sub> importer carried a flux of 8.00 mmol/gDW/h was the lowest possible and it did not

decrease any further when the system was allowed to utilise more O<sub>2</sub>, due to reactions involved in aerobic respiration reaching a saturation point.

### Responsive reactions (changers)

Between 10 and 15 reactions (Table 6-1) were required for ATP production in the 4 single solutions corresponding to the constraints applied in the extreme points of the O<sub>2</sub> scans, and between 17 and 19 reactions were active throughout the whole range of conditions considered in these scans (Table 6-4). From those, a total of 16 to 18 reactions responded to the variation in O<sub>2</sub> with a total change in flux above a threshold of 0.01 mmol/gDW/h. A total of 13 reactions responded independently of the presence or absence of NO<sub>3</sub><sup>-</sup>. These corresponded to the transporters for O<sub>2</sub>, Glt, H<sub>2</sub>O, NH<sub>4</sub><sup>+</sup>, CO<sub>2</sub> and Suc plus those reactions involved in the catabolism of Glt to Suc, aerobic respiration and the ATP synthase. The following reactions responded depending on the presence or absence of NO<sub>3</sub><sup>-</sup>: when NO<sub>3</sub><sup>-</sup> was present, the NO<sub>3</sub><sup>-</sup> importer and the reactions involved in anaerobic respiration (cytochrome-c NO<sub>2</sub><sup>-</sup> reductase and quinol-NO<sub>3</sub><sup>-</sup> reductase) became responsive. Without NO<sub>3</sub><sup>-</sup>, those 3 reactions became inactive and 5 extra reactions changed flux, corresponding to the uptake and catabolism of citrulline, including the Pro exporter.

All reactions changing flux are included in Figure 6-1 to Figure 6-3. All reactions shown in these diagrams responded to the O<sub>2</sub> variation, with the exception of the ATPase reaction, which flux was fixed as a constraint of the analysis in order to force the system to meet the NGAM and GAM ATP demand of the organism.

### Elementary modes

The net stoichiometries corresponding to the elementary modes obtained from the submodels constructed as described in Section 6.2.1.3, hence derived from the reactions active at any point of these O<sub>2</sub> scans are shown below and are classified depending on the electron acceptors utilised:

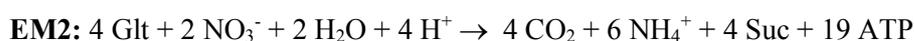
#### Net stoichiometry and ratios of modes with O<sub>2</sub> as final electron acceptor



$$\text{P/O} = 2.75$$

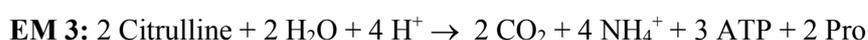
$$\text{ATP/C} = 1.1 ; \text{ATP/Glt} = 5.5$$

#### Net stoichiometry and ratios of modes utilising NO<sub>3</sub><sup>-</sup> as final electron acceptor



$$\text{ATP/C} = 0.95 ; \text{ATP/Glt} = 4.75$$

#### Net stoichiometry and ratios of modes utilising neither O<sub>2</sub> nor NO<sub>3</sub><sup>-</sup>



$$\text{ATP/C} = 0.25 ; \text{ATP/citrulline} = 1.5$$

These modes summarise the net stoichiometry of the 4 solutions corresponding to the extreme points of the scans: while modes corresponding to EM1 and EM3 constitute the two possible minimal routes for ATP synthesis occurring in the absence of  $\text{NO}_3^-$  and depending on whether  $\text{O}_2$  is present or absent, modes corresponding to EM1 and EM2 represent the two alternative routes for ATP synthesis depending on the presence of  $\text{O}_2$  when  $\text{NO}_3^-$  is available. Interestingly, the production of Pro as a result of catabolising citrulline for ATP synthesis could be a factor contributing to the apparent growth delay described in the absence of Pro in previous chapters: since RP62A was originally isolated from a catheter infection (and presumably from a biofilm), it is plausible that these cells were adapted to utilise citrulline in this manner, hence obtaining Pro without the need to activate or produce enzymes involved in its synthesis from other amino acids.

### **6.3.2 Production of PIA**

The results obtained applying the methods defined in Section 6.2 to the study of PIA production in joints are described below. Note that in order to fully understand which reactions are purely involved in the synthesis of this polymer, production of ATP was not considered in these analysis and, therefore, no flux constraint was applied to the ATPase reaction.

#### ***6.3.2.1 Metabolic responses for production of PIA in synovial fluid in the presence and absence of $\text{O}_2$ and $\text{NO}_3^-$***

Single solutions for production of PIA under conditions reproduced by the constraints applied in the extreme points of the  $\text{O}_2$  scans, thus under either total availability or absence of  $\text{O}_2$  and  $\text{NO}_3^-$ , were computed and the results obtained are presented in this section (Figure 6-6 to Figure 6-9). Greyed out reactions in these diagrams did not carry flux.

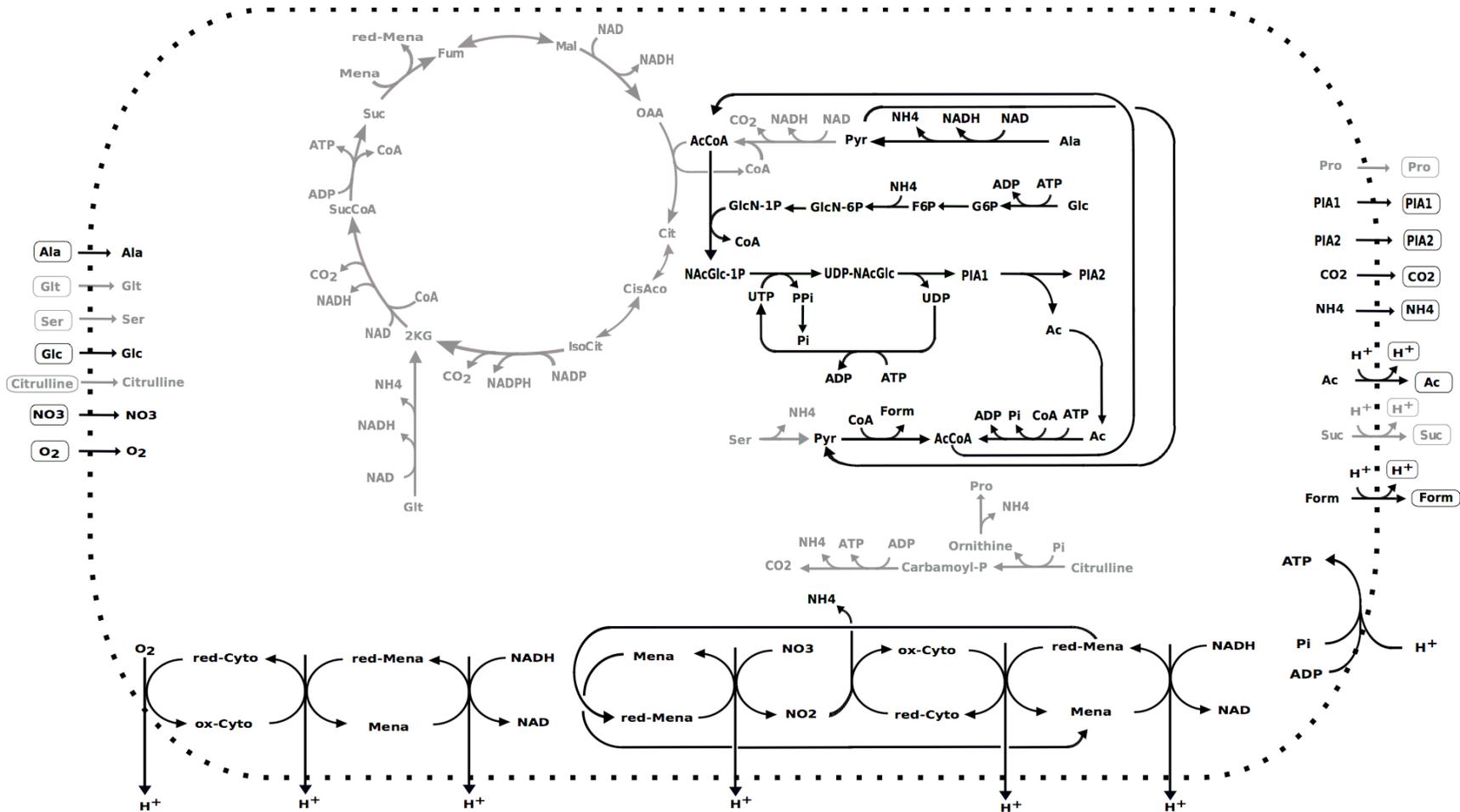


Figure 6-6 Production of PIA in synovial fluid in the presence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> when total flux through the system was minimised.

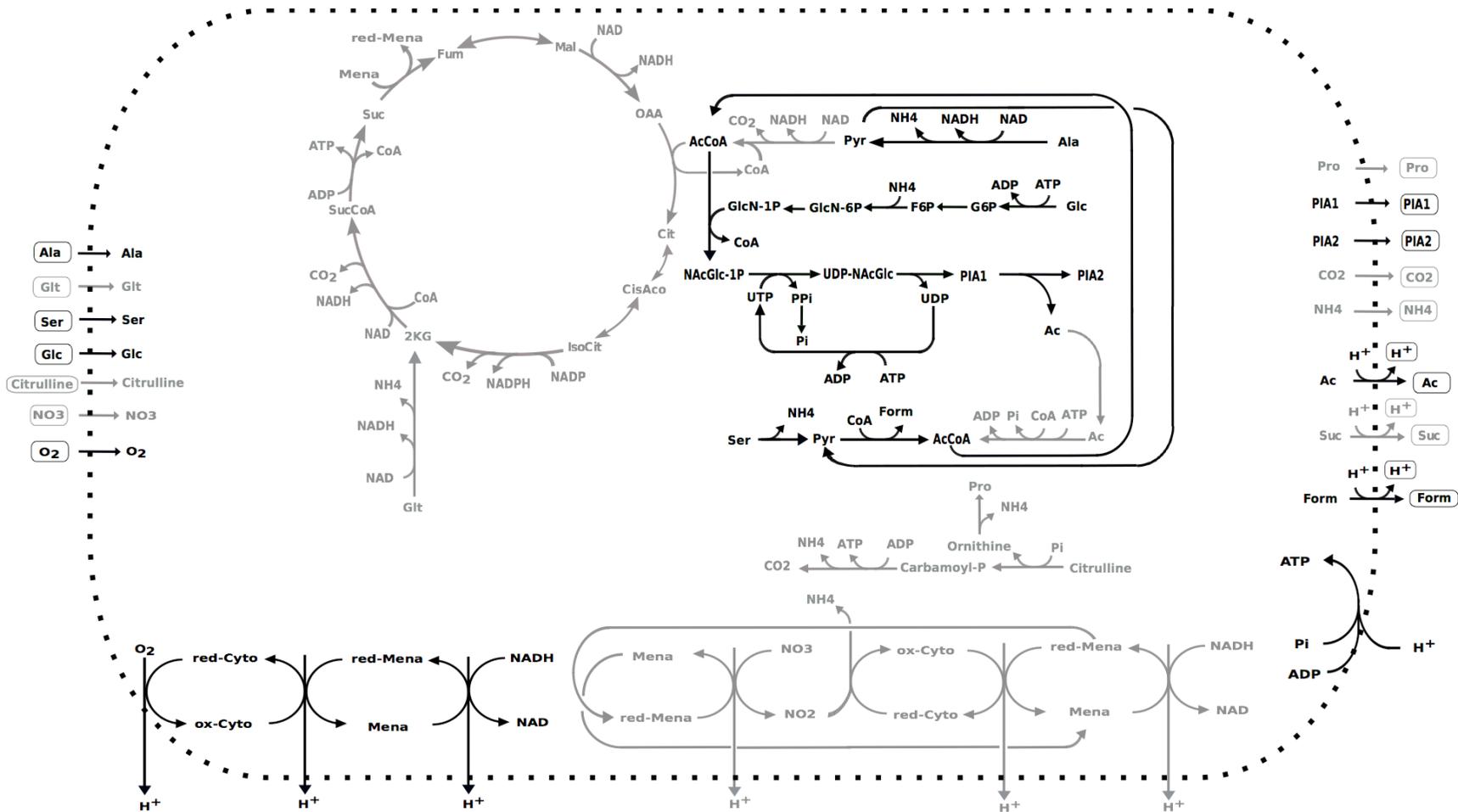


Figure 6-7 Production of PIA in synovial fluid in the presence of O<sub>2</sub> and absence of NO<sub>3</sub><sup>-</sup> when total flux through the system was minimised.

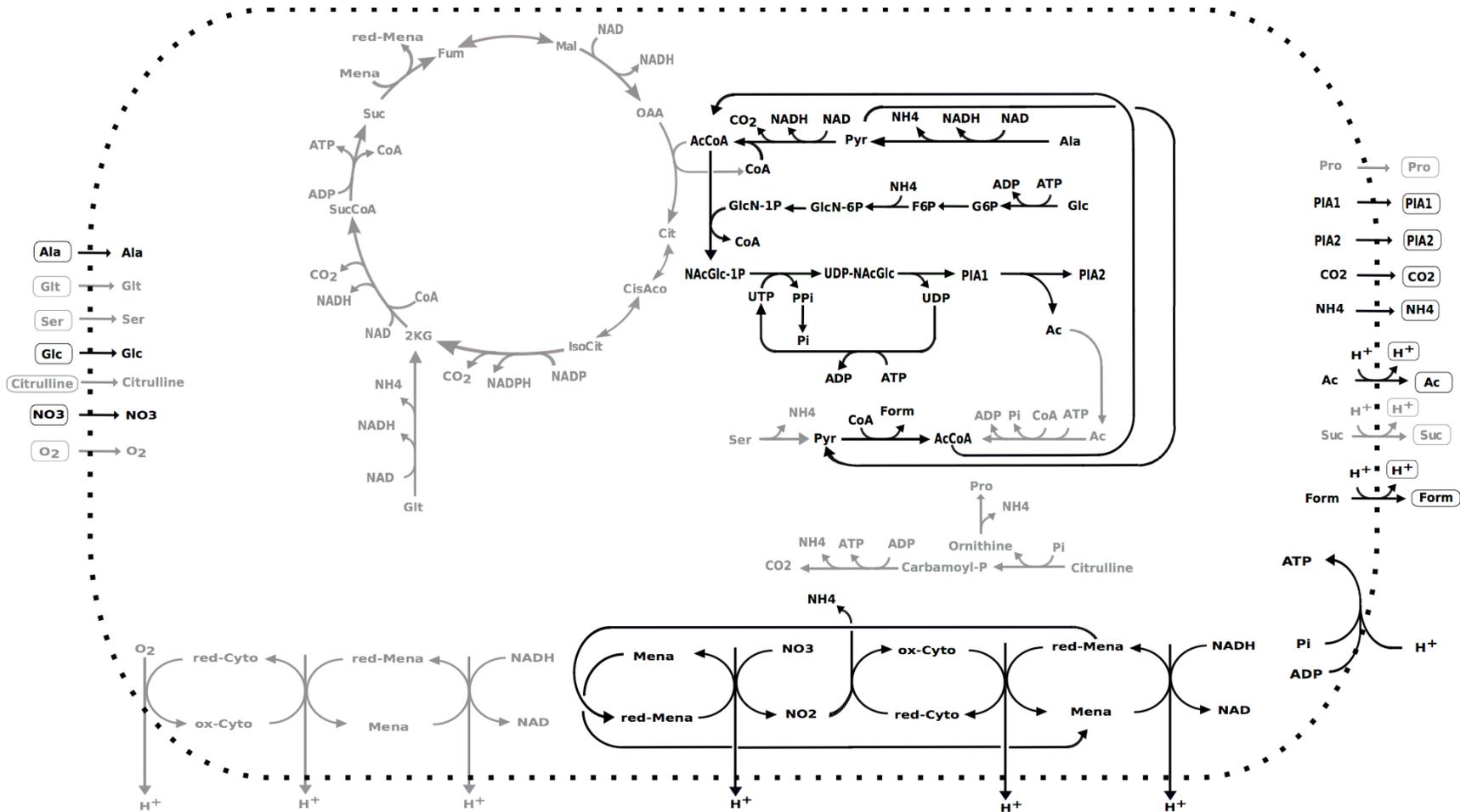


Figure 6-8 Production of PIA in synovial fluid in the presence of NO<sub>3</sub><sup>-</sup> and absence of O<sub>2</sub> when total flux through the system was minimised.

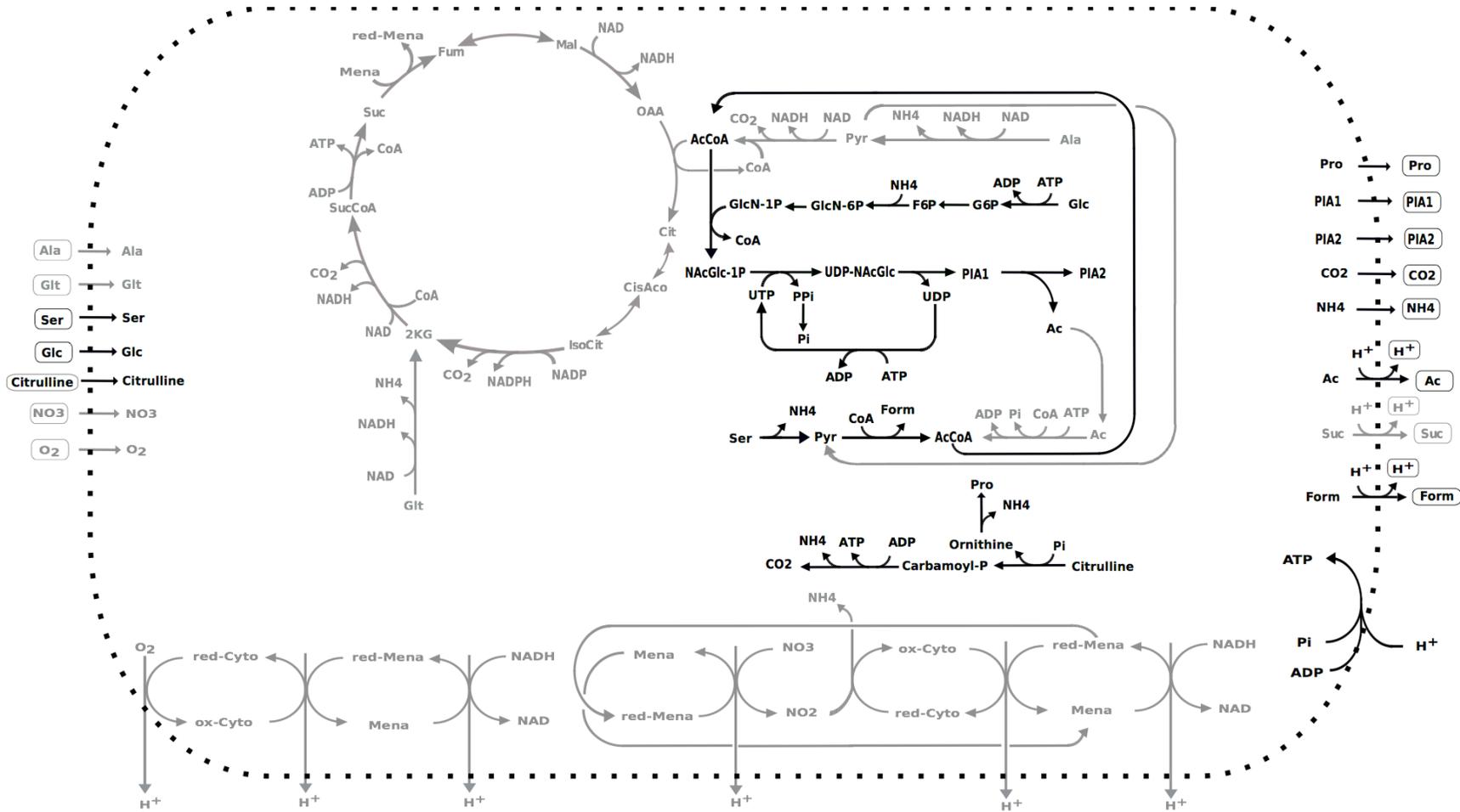


Figure 6-9 Production of PIA in synovial fluid in the absence of electron acceptor when total flux through the system was minimised.

During production of PIA in synovial fluid in the presence of  $O_2$  and  $NO_3^-$  when total flux through the system was minimised (Figure 6-6) Ala was degraded to Pyr by the Ala-dh (EC 1.4.1.1), generating NADH to be fed into the ETC, which was then fermented to Form by the Pyr Form-lyase (EC 2.3.1.54) while producing AcCoA. Glc was used to obtain GlcN-1P. AcCoA and GlcN-1P were consumed in the production of NAcGlc-1P and subsequently UDP-NAcGlc and PIA1, which was then partly deacetylated to PIA2, generating Ac. Interestingly, part of the Ac generated was used to produce more AcCoA by the Ac-CoA ligase (EC 6.2.1.1), helping to minimise total flux through the system. Functioning of the ETC via aerobic and anaerobic respiration, together with a small contribution from the excretion of both Ac and Form, generated the PMF needed for the ATP synthase to cover the ATP demand of the process.

When  $O_2$  was the only electron acceptor available, a similar response was observed (Figure 6-7), but in this case there was no flux through reactions involved in anaerobic respiration and a small amount of Ser was degraded to produce more Pyr for AcCoA synthesis. All the Ac generated by deacetylation of PIA 1 was excreted out of the system instead of used to obtain more AcCoA.

When  $NO_3^-$  was the only electron acceptor available (Figure 6-8), part of the Pyr obtained from degradation of Ala was decarboxylated leading to production of AcCoA and generating more NADH to be feed into the ETC. Glc was utilised in the same manner as in the presence of  $O_2$ , however, Ser uptake and catabolism was absent in this response. All the Ac generated by decarboxylation of PIA 1 was excreted out of the system, contributing to generate the PMF that pumps the ATP synthase.

In absence of  $O_2$  and  $NO_3^-$  (Figure 6-9), Ala was no longer utilised. Glc was imported and metabolised for production of GlcN-1P in the same manner as in the presence of electron acceptors. Ser was deaminated producing Pyr, which was further metabolised via the Pyr Form-lyase (EC 2.3.1.54) in the sole AcCoA-forming reaction of this solution, generating Form as a by-product. Citrulline was taken up and metabolised in a similar fashion as previously described for ATP synthesis under similar conditions (Section 6.3.1.1, Figure 6-3), finally leading to production and excretion of  $CO_2$ ,  $NH_4^+$  and Pro, generating ATP in the process. The excretion of charged by-products (Ac and Form) contributed to create the PMF needed for functioning of the ATP synthase, generating more ATP.

The main parameters defining these responses are summarised in Table 6-5 and Table 6-6:

**Table 6-5 Characterisation of the model behaviour for production of PIA in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.**

Electron acceptor available	Reactions carrying flux (n)	Objective value	O <sub>2</sub> uptake	NO <sub>3</sub> <sup>-</sup> uptake	Total C uptake	Total N uptake
O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup>	28	99.9	2.22	0.058	42.4	4.73
O <sub>2</sub>	26	99.9	2.29	0.00	42.5	4.73
NO <sub>3</sub> <sup>-</sup>	28	101	0.00	1.41	42.5	6.14
None	27	128	0.00	0.00	79.2	23.1

Units of the objective value and uptake rates: mmol/gDW/h. Conditions set: presence of O<sub>2</sub>, absence of O<sub>2</sub> and presence and absence of NO<sub>3</sub><sup>-</sup>.

**Table 6-6 Main N and C sources utilised and by-products excreted during production of PIA in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.**

Electron acceptor available	Compound utilized	Contribution to total N uptake (%)	Contribution to total C uptake (%)	By-products excreted
O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup>	Ala	98.8	33.1	Ac, Form
	NO <sub>3</sub> <sup>-</sup>	1.24	-	
	Glc	-	66.9	
O <sub>2</sub>	Ala	97.1	32.3	Ac, Form
	Ser	2.94	0.929	
	Glc	-	66.7	
NO <sub>3</sub> <sup>-</sup>	Ala	77.0	33.3	Ac, CO <sub>2</sub> , Form
	NO <sub>3</sub> <sup>-</sup>	23.0	-	
	Glc	-	66.7	
None	Citrulline	79.5	46.3	Ac, CO <sub>2</sub> , Form, Pro
	Ser	20.5	17.9	
	Glc	-	35.8	

As expected, PIA production was achieved with a lower enzymatic cost by utilising O<sub>2</sub> as the final electron acceptor. However, when both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> were available, a very small amount of NO<sub>3</sub><sup>-</sup> was taken up, allowing functioning of reactions involved in aerobic and anaerobic respiration simultaneously, although this last ones at a very low rate (Table 6-5 and Figure 6-6). When O<sub>2</sub> was the only electron acceptor it was utilised almost as efficiently, with a difference in the objective value

of both solutions lower than 0.01 mmol/gDW/h. PIA production in sole presence of  $\text{NO}_3^-$  or in absence of both  $\text{O}_2$  and  $\text{NO}_3^-$  was less efficient.

When  $\text{NO}_3^-$  was available, with or without  $\text{O}_2$ , Glc and Ala were the main C sources utilised. The response obtained in the sole presence of  $\text{O}_2$  was very similar, although with a small contribution of Ser to the total C uptake (< 1.00%). In the absence of both, citrulline, Glc and Ser were consumed, with Ser utilisation increasing significantly under these circumstances (by 16.9 %).

The ability of the system to produce PIA upon depletion of all free amino acids and urea was tested (Table 6-7), showing that, although presenting lower C and N uptake rates, synthesising PIA from Glc and  $\text{NH}_4^+$  increased the total net flux through the system by 32.0%.

**Table 6-7 Characterisation of the model behaviour for production of PIA in the absence of  $\text{O}_2$  and  $\text{NO}_3^-$  with Glc and  $\text{NH}_4^+$  as sole C and N sources when total net flux through the network was minimised.**

Compounds generated	Reactions carrying flux (n)	Objective value	Increase in objective value (%)	Total C uptake	Total N uptake
PIA	37	169	32.0	53.7	4.73

Units of the objective value and uptake rates: mmol/gDW/h. The increase in the objective value shown here refers to the value obtained in standard synovial fluid.

### ***6.3.2.2 Simulated variation in the $\text{O}_2$ concentration during production of PIA: responsive reactions, subnetworks and elementary modes***

After identifying single solutions for the synthesis of PIA under the most extreme conditions encountered by cells in joints, the system was analysed for PIA production under constraints that mimicked the  $\text{O}_2$  gradient found across biofilm layers following the method described in Section 6.2.1.2. This allowed identifying coordinated responses for PIA synthesis in response to changes in the  $\text{O}_2$  level. The results obtained are presented in this section. The main parameters defining these responses are summarised in Table 6-8:

**Table 6-8 Summary of data obtained in the O<sub>2</sub> scans for PIA production in synovial fluid when total net flux through the network was minimised in the presence and absence of NO<sub>3</sub><sup>-</sup>.**

NO <sub>3</sub> <sup>-</sup> available	Reactions active through the scans (n)	Changers (n)	Transporters in changers	EMs in submodel (n)
Yes	31	18	Ala, O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup> , H <sub>2</sub> O, NH <sub>4</sub> <sup>+</sup> , CO <sub>2</sub> , Ac, Form	17
No	46	33	Ala, citrulline, Glc, Glt, Ser, O <sub>2</sub> , H <sub>2</sub> O, NH <sub>4</sub> <sup>+</sup> , CO <sub>2</sub> , Suc, Pro	16

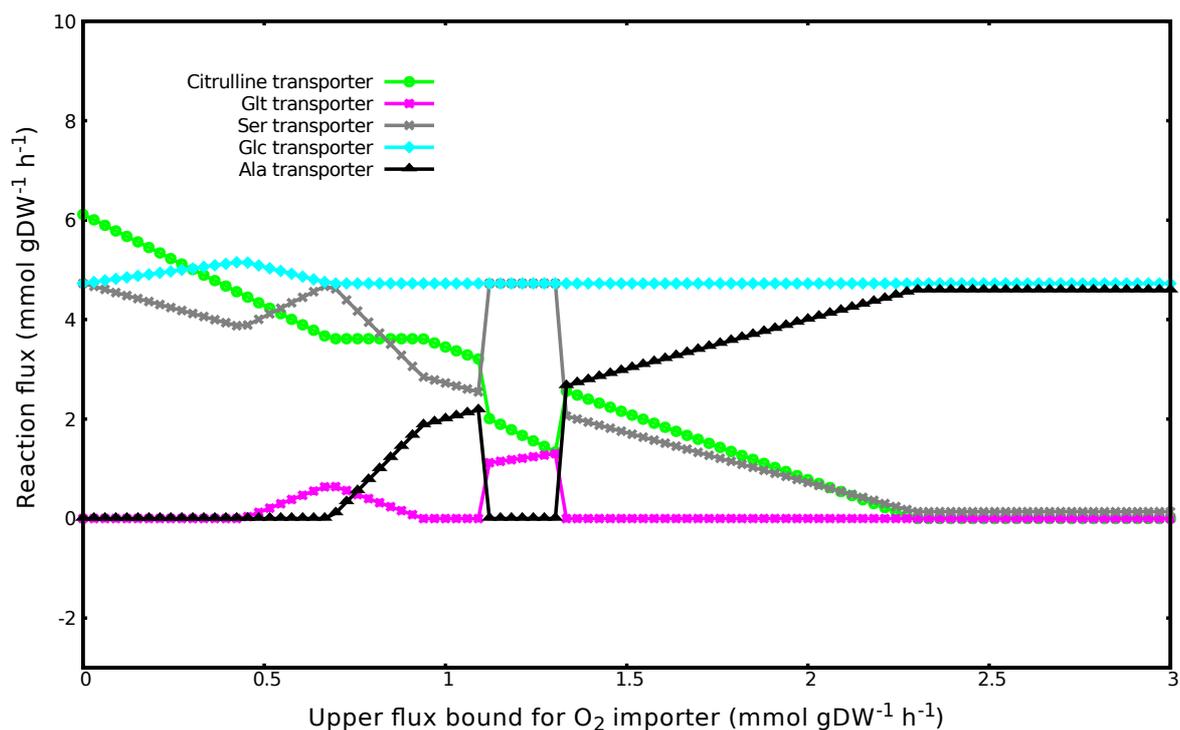
Changers = reactions with a total change in flux above a threshold of 0.01 mmol (/gDW/h)

### Responsive reactions (changers)

Between 28 and 26 reactions (**Error! Reference source not found.**) were required for PIA production in the 4 single solutions corresponding to the constraints applied in the extreme points of the O<sub>2</sub> scans, and between 31 and 46 reactions were active throughout the whole range of conditions considered in the scans. From those, a total of 18 and 33 reactions responded to the simulated variation in O<sub>2</sub> concentration with a change in flux above a threshold of 0.01 mmol (/gDW/h) and total of 10 reactions responded independently of the presence or absence of NO<sub>3</sub><sup>-</sup>. These corresponded to the transporters for Ala, O<sub>2</sub>, H<sub>2</sub>O, NH<sub>4</sub><sup>+</sup>, CO<sub>2</sub>, the Ala-dh reaction, reactions involved in aerobic respiration and the ATP synthase.

The following reactions responded to the O<sub>2</sub> variation depending on the presence or absence of NO<sub>3</sub><sup>-</sup>: when this was present, the NO<sub>3</sub><sup>-</sup> importer and the reactions involved in anaerobic respiration (cytochrome-c NO<sub>2</sub><sup>-</sup> reductase and quinol-NO<sub>3</sub><sup>-</sup> reductase) became responsive, and so did reactions leading to synthesis of AcCoA from Pyr and Ac, and the exporters for Ac, Form and CO<sub>2</sub>. Without NO<sub>3</sub><sup>-</sup>, those 8 reactions became inactive and 15 other reactions changed flux, corresponding to the uptake and catabolism of citrulline, Ser, Glt and the whole glycolysis pathway from Glc to Pyr. Some but not all of the reactions changing flux are included in Figure 6-6 to Figure 6-9, corresponding to single solutions obtained in the extreme points of the scans (e.g. reactions involved in Glt degradation or glycolysis are absent). From the reactions shown in these diagrams, those leading to synthesis of PIA1 and PIA2 from AcCoA and GlcN-1P, and including synthesis of GlcN-1P from F6P, maintained a constant flux across the scans in order to comply with the fixed PIA demand.

In order to facilitate the investigation of model responses to the changing O<sub>2</sub> levels when NO<sub>3</sub><sup>-</sup> was absent, and therefore, between conditions that allowed aerobic respiration or forced the system to exert a fermentative behaviour, the flux through importers of medium components across the O<sub>2</sub> scan were plotted in Figure 6-10:



**Figure 6-10 Change in flux for responsive importers of medium components involved in PIA synthesis in synovial fluid over variation in the availability of O<sub>2</sub> in the absence of NO<sub>3</sub><sup>-</sup> when total net flux through the network was minimised.**

The values corresponding to the upper flux bound for the O<sub>2</sub> importer are plotted in the X axis while reaction fluxes are plotted in the Y axis. Units: mmol/gDW/h

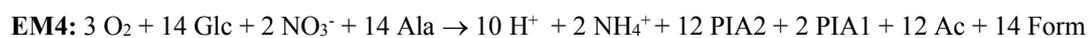
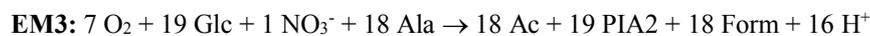
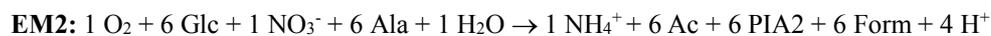
These data showed how the uptake of Glc was slightly higher at low O<sub>2</sub> concentrations, allowing production of Pyr for AcCoA synthesis via glycolysis, and remaining constant after that in order to fulfil the GlcN-1P demand. Ser degradation to Pyr was maximum at lower O<sub>2</sub> levels, and its contribution decreased in a mirror-like fashion with the activity of the glycolytic reactions, and later on, decreasing with the increasing degradation of Ala to Pyr, although small amounts of Ser were still consumed at the higher O<sub>2</sub> levels. Citruiline degradation for ATP production was maximum at the lowest O<sub>2</sub> level and it was progressively substituted by reactions involved in aerobic respiration, carrying no flux when the availability of O<sub>2</sub> was higher. There were several intermediate solutions (corresponding to upper flux bounds for the O<sub>2</sub> importer ranging from approximately 0.50 to 1.00 mmol (/gDW/h), where a small amount of Glt was taken up and degraded to Suc in order to obtain ATP, thus decreasing citruiline utilisation at an equivalent proportion in these responses. Finally, the changes in flux observed when the upper flux bound for the O<sub>2</sub> imported ranged between 1.20 mmol (/gDW/h) and 1.29 mmol (/gDW/h) represented alternative optimal solutions were, unlike in the majority of the dataset, Glt and Ser were respectively utilised to obtain ATP and AcCoA, partly substituting consumption of citruiline and completely stopping the uptake of Ala. If the system was analysed again, now blocking the flux through the Glt importer, new solutions were obtained in which the consumption of Ser decreased and utilisation of citruiline and Ala increased to levels in

line with those presented in solutions where the import of O<sub>2</sub> was outside of this range (1.20 – 1.29 mmol/gDW/h). These exhibited identical objective values as the original solutions for the exact same constraints (except for the block of flux through the Glt importer), indicating that are equally optimal for the same analysis objective (producing PIA while minimising total net flux through the network). Therefore, this exemplifies the existence of multiple optimal solutions for a given LP problem, as was previously mentioned in Section 2.2.4. However, in the interest of time and since solutions obtained at O<sub>2</sub> levels around those values consistently involved utilisation of Ala for AcCoA synthesis, a more detailed investigation of alternative optimal solutions was considered out of the scope of this study.

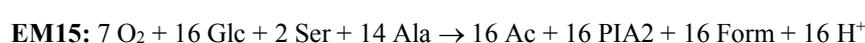
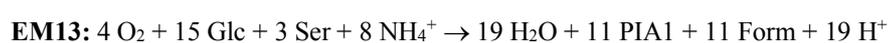
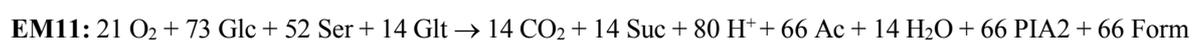
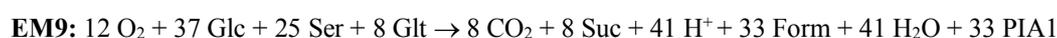
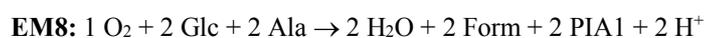
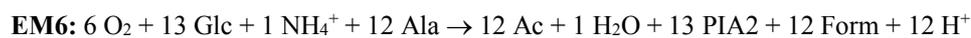
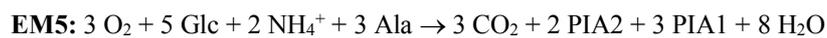
### Elementary modes

The net stoichiometries corresponding to the elementary modes obtained from the submodels constructed as described in Section 6.2.1.3, hence derived from the reactions active at any point of the O<sub>2</sub> scans for PIA synthesis, are shown below and are classified depending on the electron acceptors utilised. Each elementary mode represents an independent route through the network:

#### Net stoichiometry of modes utilising O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> as final electron acceptors

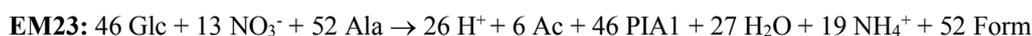
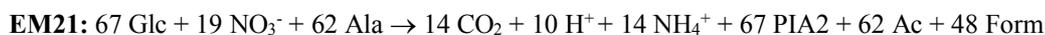
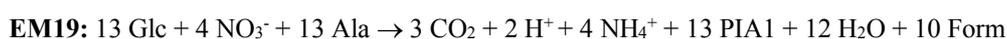
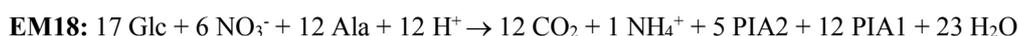


#### Net stoichiometry of modes utilising O<sub>2</sub> as final electron acceptor

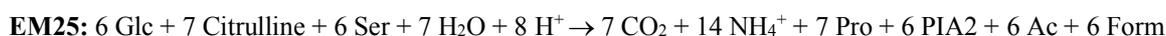


#### Net stoichiometry of modes utilising NO<sub>3</sub><sup>-</sup> as final electron acceptor

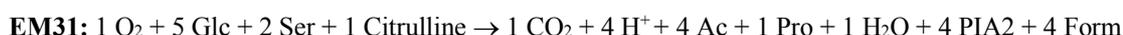
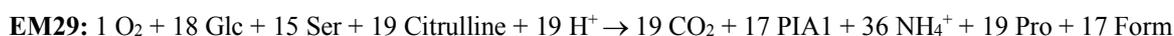
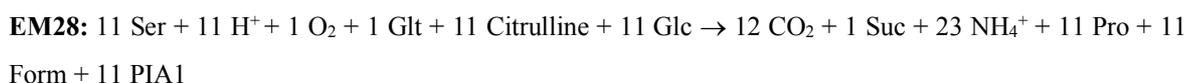
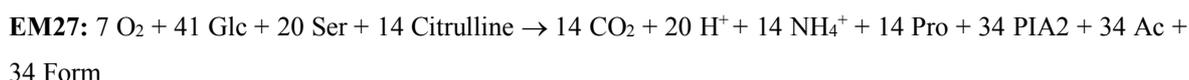




#### Net stoichiometry of modes utilising neither O<sub>2</sub> nor NO<sub>3</sub><sup>-</sup> as final electron acceptors



#### Net stoichiometry of modes utilising O<sub>2</sub> as final electron acceptor involving citrulline degradation



The large number of modes obtained denotes the high versatility of these subnetworks, and therefore, of the system they derive from. Notice that, while different mechanisms can lead to ATP and AcCoA production, the only plausible way to obtain GlcN-1P involves metabolising Glc. Hence, Glc is consumed in every single mode, while the utilisation of several substrates, such as Ala, Ser, citrulline, Glt and Glc can be combined in different ways in order to obtain ATP and AcCoA.

### 6.3.3 Production of planktonic biomass

This section describes the results obtained applying the methods presented in Section 6.2 to the study of the synthesis of planktonic biomass under the environmental conditions found in joints. Since the objective of this work is to fully understand which processes are relevant for either ATP synthesis, production of PIA or biosynthesis of planktonic biomass, no flux constraint was applied to the ATPase reaction during the analyses performed for biomass production.

### 6.3.3.1 *Metabolic responses for production of planktonic biomass in synovial fluid in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>*

Single solutions for production of planktonic biomass obtained applying the constraints used at the extreme points of the O<sub>2</sub> scans, thus under total availability or absence of O<sub>2</sub> and in the presence or absence of NO<sub>3</sub><sup>-</sup> have been computed and the results obtained are presented in this section. The main parameters defining these responses are summarised in Table 6-9 and Table 6-10.

**Table 6-9 Characterisation of the model behaviour for production of planktonic biomass in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.**

Electron acceptor available	Reactions carrying flux	Objective value	O <sub>2</sub> uptake	NO <sub>3</sub> <sup>-</sup> uptake	Total C uptake	Total N uptake
O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup>	287	71.0	1.49	0.00	40.4	9.96
O <sub>2</sub>	290	71.0	1.48	0.00	40.3	9.77
NO <sub>3</sub> <sup>-</sup>	290	72.0	0.00	0.856	41.4	10.86
None	287	85.0	0.00	0.00	63.4	21.0

Units of the objective value and uptake rates: mmol/gDW/h. Conditions set: presence of O<sub>2</sub>, absence of O<sub>2</sub> and presence and absence of NO<sub>3</sub><sup>-</sup>

Planktonic biomass was produced with a lower associated total net flux through the system by utilising O<sub>2</sub> as the final electron acceptor. When both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> were available, only O<sub>2</sub> was utilised. The responses obtained in sole presence of NO<sub>3</sub><sup>-</sup> or in the absence of both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> were less efficient.

**Table 6-10 Main N and C sources utilised and by-products excreted during production of planktonic biomass in synovial fluid when total net flux through the network was minimised in the presence and absence of electron acceptors.**

Electron acceptor available	Compound utilized	Contribution to total N uptake (%)	Contribution to total C uptake (%)	By-products excreted
<b>O<sub>2</sub>, NO<sub>3</sub><sup>-</sup></b>	Gln	20.6	12.6	CO <sub>2</sub> , Form, Suc
	Ala	20.1	14.8	
	Thr	8.66	8.53	
	Asp	7.61	7.49	
	Lys	6.70	5.02	
	Glc	-	16.1	
<b>O<sub>2</sub></b>	Ala	20.4	14.9	CO <sub>2</sub> , Form, Suc
	Gln	19.7	12.0	
	Thr	8.83	8.57	
	Asp	6.55	6.36	
	Lys	6.92	5.04	
	Glt	5.59	6.78	
	Glc	-	17.1	
<b>NO<sub>3</sub><sup>-</sup></b>	Ala	18.4	14.5	CO <sub>2</sub> , Form, Suc
	Gln	17.8	11.6	
	Thr	7.94	8.33	
	NO <sub>3</sub> <sup>-</sup>	7.88	-	
	Glt	7.18	9.40	
	Glc	-	16.6	
<b>None</b>	Citrulline	56.7	37.6	CO <sub>2</sub> , Form, Suc, Pro
	Ser	10.7	10.6	
	Glt	5.17	8.58	
	Gln	4.80	0.506	
	Thr	3.44	4.57	
	Glc	-	10.2	

When several N and/or C sources were taken up, data corresponding to at least the five compounds contributing to these total uptakes in the highest proportions were included in the table.

In presence of O<sub>2</sub>, Ala and Gln were the main N sources consumed, followed by Thr, Asp and Lys etc. This compares well with the amino acids defined as contributing in the higher proportion to the N and C demand for planktonic growth in Chapter 5, Section 5.3.1.2, where Glt, Ala, Arg, Thr and Lys were shown to account for the majority of the N demand, and Glt, Ala, Thr, Asp and Lys for the

C demand. When only  $\text{NO}_3^-$  was available, the contribution of Glt increased over that of Asp and Lys. In the absence of both electron acceptors, citrulline and Ser became the main N sources utilised, followed by Glt, Gln and Thr. This led to Pro being produced and excreted out of the system.

The model was also capable of synthesising biomass under highly restrictive conditions (absence of amino acids and electron acceptors), and the main features of this response are shown in Table 6-11. However, the dramatic increase in the total net flux associated with this response when compare to growth in synovial fluid (+227%), suggests that planktonic growth could be reduced or impaired under these conditions *in vivo*, for example, in cells living in the bottom layers of a biofilm.

**Table 6-11 Characterisation of the model behaviour for production of planktonic biomass in the absence of  $\text{O}_2$  and  $\text{NO}_3^-$  with Glc and  $\text{NH}_4^+$  as sole C and N sources when total net flux through the network was minimised.**

Compounds generated	Reactions carrying flux	Objective value	Increase in objective value (%)	Total C uptake	Total N uptake
Planktonic biomass	338	278	227	60.7	7.25

Units of the objective value and uptake rates: mmol/gDW/h. The increase in the objective value shown here refers to the value obtained in standard synovial fluid.

### ***6.3.3.2 Simulated variation in the $\text{O}_2$ concentration during production of planktonic biomass: responsive reactions, subnetworks and elementary modes***

The system was analysed for production of planktonic biomass applying constraints that mimicked the  $\text{O}_2$  gradient found across biofilm layers following the method described in Section 6.2.1.2. The transporters which responded to the variation in the  $\text{O}_2$  level are reported in Table 6-12. These give an indication towards which areas of metabolism respond to changing conditions during synthesis of planktonic biomass, and could potentially be confirmed experimentally in the future. An overview of the reactions responding depending on the presence or absence of  $\text{NO}_3^-$  is given below.

**Table 6-12 Summary of data obtained with O<sub>2</sub> scans for production of planktonic biomass in synovial fluid and the presence and absence of NO<sub>3</sub><sup>-</sup> when total net flux through the network was minimised.**

NO <sub>3</sub> <sup>-</sup> available	Reactions active through the scans (n)	Changers (n)	Transporters in changers	EMs in submodels (n)
Yes	294	23	Glt, O <sub>2</sub> , NO <sub>3</sub> <sup>-</sup> , H <sub>2</sub> O, NH <sub>4</sub> <sup>+</sup> , CO <sub>2</sub> , Form, Suc	Non-computed
No	306	69	Ala, Arg, Citrulline, Glc, Glt, Ser, Thr, O <sub>2</sub> , H <sub>2</sub> O, NH <sub>4</sub> <sup>+</sup> , CO <sub>2</sub> , Form, Suc, Pro	Non-computed

Changers = reactions with a total change in flux above a threshold of 0.01 mmol (/gDW/h)

### Responsive reactions (changers)

Between 290 and 287 reactions (Table 6-9) were required for planktonic biomass production under the conditions corresponding to the constraints applied at the extremes of both scans. 294 and 306 reactions were active throughout the whole range of conditions considered in the scans. From those, a total of 23 to 69 reactions responded to the variation in the availability of O<sub>2</sub> with a change in flux above a threshold of 0.01 mmol (/gDW/h).

20 reactions responded to the O<sub>2</sub> variation in both O<sub>2</sub> scans, the one performed considering NO<sub>3</sub><sup>-</sup> as available and the one performed constraining its import to zero: these corresponded to transporters for Glt, O<sub>2</sub>, H<sub>2</sub>O, NH<sub>4</sub><sup>+</sup>, CO<sub>2</sub>, Form and Suc, reactions involved in aerobic respiration, and those catalysed by the ATP synthase, an Ala-dh (producing Pyr and NADH), an Ala-aminotransferase (transforming Glt and Pyr to Ala and 2-KG), the NADP-dependant Glt-dh (regenerating NADPH and producing 2-KG), the 2-KG-dh and Suc-CoA-synthase from the TCA cycle (generating Suc, NADH and ATP), two reactions from the mevalonate biosynthetic and degradation pathways (acting as a transhydrogenase system (EC 1.1.1.34 and EC 1.1.1.88), which re-oxidised excess of NADPH and generated NADH), and the Pyr-dh and Pyr Form-lyase (producing AcCoA from Pyr). Therefore, this indicates that these reactions are important for environmental adaptation regardless of whether the system still able to perform anaerobic respiration or is forced to produce fermentative responses. The following reactions responded to the O<sub>2</sub> variation depending the availability of NO<sub>3</sub><sup>-</sup>: when NO<sub>3</sub><sup>-</sup> was available, the NO<sub>3</sub><sup>-</sup> importer and the reactions involved in anaerobic respiration (cytochrome-c NO<sub>2</sub><sup>-</sup> reductase and quinol-NO<sub>3</sub><sup>-</sup> reductase) became responsive. When NO<sub>3</sub><sup>-</sup> was unavailable, those 3 reactions remained inactive and 46 new reactions changed flux, corresponding to the uptake of several amino acids (Ala, Arg, Ser, Thr and citrulline), the excretion of Pro, the degradation of citrulline and Ser (producing ATP and Pyr respectively), several dehydrogenases, deaminase reactions and those involved in the complete PPP (within others). Most of these processes involved production of NADPH and precursor metabolites (e.g. ribose-5-P etc.). 237 reactions maintained a

constant flux across both scans, complying with the fixed biomass demand independently of the environmental conditions.

## **6.4 Discussion**

### **6.4.1 ATP production in joints**

#### ***6.4.1.1 Metabolic responses for ATP production in synovial fluid in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>***

In summary, in the presence of O<sub>2</sub>, the system took up Glt and metabolised it to 2-KG generating NADH. 2-KG was further metabolised via reactions of the TCA cycle, producing ATP and NADH to be fed into the aerobic ETC while excreting Suc. When only NO<sub>3</sub><sup>-</sup> was available, this acted as the final electron acceptor in a similar solution that involved the anaerobic ETC instead. In the absence of both electron acceptors, citrulline was catabolised to carbamoyl-P and ornithine, being these compounds further degraded to CO<sub>2</sub>, Pro and NH<sub>4</sub><sup>+</sup>, while generating ATP in a process that, to our knowledge has not been described before as important for energy production in staphylococcal cells living in joints. Interestingly, upon citrulline deprivation, the system did not import Arg in order to obtain citrulline for ATP synthesis via Arg deamination (EC 3.5.3.6), which has been shown to be important (although not essential) for biofilm formation in Staphylococci (Zhu *et al.* 2007) (Chapter 0, Section 1.6.3). Instead, Ser was deaminated to Pyr, which was further catabolised to Form and AcCoA, with this final compound being used to synthesise ATP while producing Ac. The excretion of charged by-products (Ac, Form, NH<sub>4</sub><sup>+</sup> and Suc) contributed to generate the PMF which allowed functioning of the ATP synthase, as was described when investigating the synthesis of ATP from several substrates in Chapter 3, Section 3.3.2.

#### ***6.4.1.2 Simulated variation in the O<sub>2</sub> concentration during ATP production: responsive reactions, subnetworks and elementary modes***

Results derived from the O<sub>2</sub> scan performed in the presence of NO<sub>3</sub><sup>-</sup> (Section 6.3.1.2), specifically highlighted the importance of the Glt uptake and metabolism for ATP synthesis in RP62A, and showed how reactions involved in aerobic and anaerobic respiration changed flux in a mirror-like fashion when responding to the variation in the availability of O<sub>2</sub>. These data showed that, even in the presence of Glc, the system preferentially consumed free amino acids for ATP production. Despite glycolysis and the TCA cycle being pathways traditionally described as leading to the greatest ATP/C yields, when other plausible biological objectives such as minimising protein cost are considered (represented here by minimisation of the total net flux through the system as a proxy), alternative catabolic routes take over. In presence of O<sub>2</sub> or NO<sub>3</sub><sup>-</sup>, Glt degradation generated ATP and NADH to be fed to the ETC, while in their absence, citrulline (or Ser) degradation allowed an optimal

balance between ATP synthesis and total net flux minimisation. In summary, this work identifies the ATP synthase reaction and reactions involved in Glt degradation, aerobic and anaerobic respiration and citrulline and Ser degradation as important for adaptation to energy production in the array of conditions found in joints, and suggests that reaction involved in aerobic and anaerobic respiration are simultaneously active throughout a wide range of O<sub>2</sub> levels.

These observations are consistent with the model behaviour described in Chapter 3, Section 3.3.2.2, where the production of ATP from Glt in the presence of O<sub>2</sub> was investigated, and, moreover, the importance of Glt catabolism for ATP synthesis defined in this chapter could constitute a plausible explanation for the drastic reduction in biofilm formation described in Chapter 4, Section 4.3.2, where Glt deprivation caused the greatest decrease in biofilm formation in RP62A cultures grown in MM medium when single amino acids were removed from the standard medium formulation at a time.

## 6.4.2 PIA production in joints

### 6.4.2.1 *Metabolic responses for production of PIA in synovial fluid in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>*

In all conditions, Glc was used to obtain the GlcN-1P consumed together with AcCoA in the synthesis of UDP-NAcGlc, which was further metabolised to PIA1, while PIA1 was partly deacetylated to PIA2, generating Ac before these polymers were finally excreted. This is an accurate representation of the biological process of PIA synthesis described for staphylococci (Chapter 0, Section 1.7.4) (Heilmann *et al.* 1996; Gerke *et al.* 1998; Lee *et al.* 2016; Somerville 2016), in which the transamination of F6P to GlcN-6P directs C to the synthesis of amino sugars and away from glycolysis (Somerville 2016), UDP-NAcGlc is produced via acetylation and activation of GlcN-1P, an initial polymer is synthesised by the polymerisation of NAcGlc donated by UDP-NAcGlc, being then partly deacetylated and finally excreted from the cell. Some particular features of the model responses obtained during this work are: i) the simultaneous functioning of aerobic and anaerobic respiration when both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> are available; ii) the utilisation of part of the Ac generated during the deacetylation of PIA1 to obtain more AcCoA; and, iii) the use of Ser to obtain AcCoA both in aerobic and anaerobic conditions. Interestingly, these coincide with the following biological observations in staphylococci: i) the expression of both cytochromes and the nitrate reductase has been detected in microaerobic conditions in *S. epidermidis* (Uribe-Alvarez *et al.* 2016), implying the simultaneous functioning of the aerobic and anaerobic version of the ETC; ii) staphylococci are known to consume Ac and use it to obtain AcCoA, specially during the post-exponential growth state (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016; Halsey *et al.* 2017); and, iii) the use of Ser to obtain ATP and AcCoA via Pyr has also been described in *S. epidermidis* both in the presence and absence of O<sub>2</sub> (Sivakanesan *et al.* 1980). The contribution of the excretion of charged by-products (Ac and Form) to generate the PMF which allows functioning of the ATP synthase in a

similar way as described during the study of ATP synthesis in Chapter 3 cannot be either verified or disproved due to the current lack of data on this matter, thus further studies are needed in order to test this hypothesis.

#### ***6.4.2.2 Simulated variation in the O<sub>2</sub> concentration during production of PIA: responsive reactions, subnetworks and elementary modes***

Results derived from the O<sub>2</sub> scan in the presence of NO<sub>3</sub><sup>-</sup> (Section 6.3.2.2) showed how reactions involved in aerobic and anaerobic respiration responded to the O<sub>2</sub> variation, together with the Ala importer and the reactions leading to AcCoA synthesis from Pyr and Ac, respectively catabolised by the Pyr Form-lyase and the Ac-CoA ligase. Results derived from the O<sub>2</sub> scan in the absence of NO<sub>3</sub><sup>-</sup> highlighted the importance of the metabolism of Glc, Ala, Glt, Ser and citrulline in the adaptation of RP62A to produce biofilm polymers upon changes in the O<sub>2</sub> concentration, with Ala being increasingly utilised to obtain AcCoA and NADH for ATP synthesis the more O<sub>2</sub> becomes available, utilisation of Glc increasing slightly towards Pyr production at low O<sub>2</sub> levels, with its consumption otherwise being kept constant to fulfil the GlcN-1P demand, Ser and citrulline utilisation for the respective production of AcCoA and ATP decreasing almost linearly as the availability of O<sub>2</sub> increased and reactions involved in aerobic respiration augmenting their flux. Glt was utilised to obtain ATP for PIA synthesis in a minority of the responses in which O<sub>2</sub> was available.

PIA production was achieved at a lower enzymatic cost by utilising O<sub>2</sub> as the final electron acceptor. When both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> were available, the most efficient response was achieved by reactions involved in both aerobic and anaerobic respiration carrying flux simultaneously, although this last ones did so at a very low rate. In presence of electron acceptors, Ala was the optimal substrate for production of the AcCoA needed for PIA synthesis. In absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>, AcCoA was synthesised via Ser degradation. GlcN-1P, also needed for PIA production, was obtained from Glc and NH<sub>4</sub><sup>+</sup> in an ATP-consuming process. This ATP was generated via aerobic and/or anaerobic respiration, or, in absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>, by substrate-level phosphorylation via citrulline degradation. Excretion of charged by-products contributed to generate the PMF which allowed functioning of the ATP synthase. Interestingly, in the absence of NO<sub>3</sub><sup>-</sup>, Glt degradation and the whole glycolysis pathway were active, and were respectively utilised to obtain ATP and Pyr at intermediate O<sub>2</sub> levels, although this occurred in a very limited number of responses. These strategies allowed PIA production with the highest possible reduction in the total net flux through the system.

In summary, this work indicates that reactions involved in Ala degradation, aerobic and anaerobic respiration, the ATP synthase, Ser and citrulline degradation, and, to a lesser extent, in glycolysis and Glt degradation are important for RP62A adaptation to PIA production under the range of conditions found across biofilms in joints. Despite synthesis of PIA being an ATP consuming process, the reactions identified during the study of ATP production in Section 6.3.1, such as the

import and degradation of Glt, carried no flux in most of these solutions. This seems to indicate that production of PIA could work as an anabolic process coupled with catabolic reactions that contribute to fulfil its associated energy demand, such as the degradation of amino acids (e.g. oxidation of Ala, which generates NADH to enter the ETC) or the deacetylation of PIA1 (generating Ac which excretion contributes to the PMF that pumps the ATP synthase), while the system optimally minimises the total net flux through the system. However, here we hypothesise that since 6% of the NAcGlc residues in PIA are O-succinated, Glt degradation to 2-KG and its metabolization to Suc, with the consequent production of ATP and NADH, are also likely to be involved in PIA synthesis, although the current mechanism of the PIA succination process remains unknown and thus has not been considered in the model. This would again suggest that Glt catabolism is important for biofilm formation, as shown by the *in vitro* results described in Chapter 4, Section 4.3.2.

According to the results derived from these analyses, pathways traditionally described as central to the production of energy and biosynthetic precursors, such as glycolysis, the TCA cycle or the PPP are either absent or have very little involvement in PIA production. Interestingly, external factors leading to repression of the TCA cycle (e.g. high Glc level, nutrient and iron deprivation and decreased O<sub>2</sub> concentrations) have been shown to dramatically increase PIA synthesis *in vitro* (Vuong *et al.* 2005; Sadykov *et al.* 2008; Sadykov *et al.* 2011; Somerville 2016). Therefore, these biological observations support the validity of our results. Again, the mathematical analysis of the metabolic network has been proven useful in anticipating regulatory patterns that allow the living organism to implement optimal strategies for achieving certain biological objectives: e.g. PIA synthesis would be optimally achieved with the lowest associated enzymatic cost via repression of the TCA cycle, utilising alternative routes for production of energy and metabolic precursors. Thus reducing protein investment is suggested here has a strong candidate to represent the real biological objective followed by cells living under these conditions.

The large number of elementary modes present in the subnetworks derived from the reactions involved in PIA production again demonstrates the versatility of the network with regards to this biosynthetic process, and, therefore suggests it has a high potential to adapt to changing environmental conditions. While different mechanisms led to production of the ATP and AcCoA consumed in the process, all identified routes generating GlcN-1P involved metabolising Glc. Hence Glc was consumed in every single mode, while several substrates, such as Ala, Ser, citrulline or Glt could be combined in different ways in order to comply with the ATP and AcCoA demand. This behaviour matches findings by other authors who reported that an increase in the Glc concentration promoted biofilm formation (Vuong *et al.* 2005; Sadykov *et al.* 2008; Sadykov *et al.* 2011; Somerville 2016). Finally, the large number of routes leading to synthesis of either PIA1, PIA2 (positively charged) or both, suggested that the system could easily control the ratio between the acetylated and deacetylated residues in PIA, thus rapidly modifying the charge of the biofilm matrix

in response to the environment and the stage of the biofilm formation process itself. This could either contribute to the adhesion of the biofilm to surfaces with different charges and to the accumulation, maturation or detachment phases of biofilm formation as per required (Chapter 0, Section 1.7.1). Interestingly, cells in biofilms have already been described to vary the relative proportion of teichoic acid and PIA with the very same purpose (Gross *et al.* 2001; Sadovskaya *et al.* 2004).

Here we have identified a set of optimal solutions for PIA production and what could potentially be a proxy for enzyme salvage in joints. However, other optimal and suboptimal solutions could exist. Therefore, a more complete study of all possible routes leading to synthesis of PIA could define single reactions unconditionally essential for this process that could then be proposed as potential new therapeutic targets. Such a detailed investigation is out of the scope of this project but will be performed in the near future as a continuation of this work.

### **6.4.3 Production of planktonic biomass in joints**

#### ***6.4.3.1 Metabolic responses for production of planktonic biomass in synovial fluid in the presence and absence of O<sub>2</sub> and NO<sub>3</sub><sup>-</sup>***

Planktonic biomass was produced at a lower enzymatic cost by utilising O<sub>2</sub> as the final electron acceptor. When O<sub>2</sub> or NO<sub>3</sub><sup>-</sup> were available, Ala and Gln were the main amino acids consumed, followed by others. In absence of both electron acceptors, citrulline and Ser took their place, leading to Pro being excreted from the system. Glc was highly consumed as a C source in all conditions. In summary, these results indicate that in the presence of electron acceptors, consumption of amino acids that could be converted to Ala and the reactions involved in Ala degradation were optimally utilised for the synthesis of ATP, Pyr and subsequently the AcCoA needed for biosynthetic processes. Glycolysis was also utilised in these solutions as a way to obtain more Pyr. In the absence of electron acceptors, degradation of citrulline and Ser became the major source of ATP and AcCoA while Glc was still utilised as the third most consumed C source.

#### ***6.4.3.2 Simulated variation in the O<sub>2</sub> concentration during production of planktonic biomass: responsive reactions, subnetworks and elementary modes***

The results derived from the O<sub>2</sub> scan performed in the presence of NO<sub>3</sub><sup>-</sup> (Section 6.3.3.2), showed flux changes in reactions involved in aerobic and anaerobic respiration. Changes in the flux of the following reactions were detected during O<sub>2</sub> scans both with and without NO<sub>3</sub><sup>-</sup>: reactions involved in metabolising Glt for production of Ala, NADH and 2-KG, the metabolism of 2-KG via reactions of the TCA cycle leading to synthesis of ATP, a transhydrogenase system exchanging NADPH for NADH, and reactions generating Pyr and subsequently AcCoA from Ala.

The fact that the metabolism of Glc and the amino acids Ala, Arg, Glt, Ser and Thr was identified as important for production of planktonic biomass in these analysis (Section 6.3.3.2, Table 6-12) is not surprising since Ala, Ser and Thr are glycogenic amino acids and staphylococci have been reported to utilise them *in vitro* for production of both AcCoA and ATP (via Pyr), leading to excretion of Ac, Form and Lac (Sivakanesan *et al.* 1980; Halsey *et al.* 2017) (Chapter 0, Section 1.6.2), while the degradation of Glt to 2-KG and its metabolism to Suc leading to production of ATP and NADH has also been documented in staphylococci (Tynecka *et al.* 1999; Halsey *et al.* 2017) as well as its role as main N donor in the biosynthesis of amino acids. Finally, Arg is known to be utilised as a source of Glt to keep fuelling these processes.

Deamination reactions occurring during amino acid catabolism generate  $\text{NH}_4^+$ , and have been proposed to potentially help counteract excessive acidification of the cell caused by production of by-products such as Ac, Form, Lac or Suc (Beenken *et al.* 2004; Resch *et al.* 2005; Resch *et al.* 2006). Therefore, here we hypothesised that staphylococcal cells might modify the ratio between consumption of Glc and  $\text{NH}_4^+$  and the uptake and catabolism of amino acids for anabolic purposes as a way to help regulating the cellular pH, which is important for adapting to the environment.

Production of ATP, PIA or planktonic biomass is still possible even in the absence of electron acceptors and all amino acids, however, the substantial increase in the total net flux associate with these responses might result in growth being impaired or considerably reduced *in vivo*.

## 6.5 Conclusion

The results presented in this chapter demonstrate that the system is capable of accurately reproducing the mechanisms currently described for PIA synthesis in staphylococci. Application of the LP-based analysis technique has allowed identification of reactions that might be important for production of energy, PIA and planktonic biomass in joints. This work shows how the metabolic network rearranges itself, varying the preferential uptake and metabolism of amino acids and Glc in response to changes in the presence of  $\text{O}_2$  and  $\text{NO}_3^-$  and in the availability of  $\text{O}_2$ . The results obtained highlighted the importance of the degradation of citrulline as a way to obtain ATP in the absence of electron acceptors, a metabolic route that, to our knowledge, has not been described before in this context.

Defining the elementary modes in the subsystems composed of reactions involved in the synthesis of ATP or PIA across an array of  $\text{O}_2$  levels identified minimal routes through these sets of reactions. These could potentially be combined in different ways, allowing the cell to prioritise production of energy and/or PIA1 and PIA2 as an adaptive response to changes in the environment. Moreover,

these data have expanded our understanding in the metabolic capabilities and the flexibility of the system's network, suggesting that the redundancy of minimal routes leading to production of PIA1 and PIA2, and hence the ability of the cell to modify the level of de-acetylated residues in the biofilm matrix, is an important feature in the production of biofilms of polysaccharidic nature.

The mathematical analysis of the metabolic network has again been proven useful in anticipating regulatory patterns leading to the implementation of optimal strategies in the living organism, in this case, showing that PIA synthesis is optimally achieved (with the lowest associated total flux) via repression of the TCA cycle and utilisation of alternative routes for production of energy and metabolic precursors, as has been described *in vitro* (Vuong *et al.* 2005; Sadykov *et al.* 2008; Sadykov *et al.* 2011; Somerville 2016). This work has not only helped explain the metabolic significance of biological observations but has also provided a metabolic explanation for biological phenotypes, such as the dramatic reduction of biofilm formation in cultures grown upon depletion of Glt, which catabolism has been defined here as important for the synthesis of ATP and has been suggested as a feasible source of both ATP, and Suc for the succination of PIA residues.

## 7 General discussion

### 7.1 Comparison of GSMs of staphylococci

This is, to our knowledge, the first highly-curated GSM produced for *S. epidermidis*. The lack of previous models for this organism limited the amount of supporting data available for the model construction and curation steps. To date, there have been several GSMs published for staphylococci, most of them based on the *S. aureus* strain N315. From those, only one (Bosi *et al.* (2016)) was published in SBML format and could, therefore, be imported and analysed for its general properties with the software ScrumPy used for this project. This analysis identified 866 reactions (more than half of the reactions in the network) as dead, and stoichiometric inconsistencies for up to 789 metabolites throughout the system. Therefore, further quality checks (e.g. mass or energy conservation) were beyond consideration. For the remaining models, data directly retrieved from the original publications were used for the purpose of model comparison (Table 7-1):

**Table 7-1 General properties of curated GSMs available for staphylococci**

Model and species modelled	Reactions	Transport reactions	Reactions with gene-association	Metabolites	Dead reactions	Unbalanced reactions	Energy/redox consistency
<i>S. epidermidis</i> (this project)	952	72	606	859	339	0	Yes
<i>S. aureus</i> Bosi <i>et al.</i> (2016)	1475	NFND	NFND	1232	866	NFND	NFND
<i>S. aureus</i> Lee <i>et al.</i> (2009)	1497	146	NFND	1431	NFND	NFND	NFND
<i>S. aureus</i> iMH551 (2005)	774	92	726	712	225	NFND	Yes
<i>S. aureus</i> iSB619 (2005)	640	84	581	571	108	3	NFND

NFND = data not found or not described in the corresponding publication

The two latest models for N315 (Lee *et al.* (2009) and Bosi *et al.* (2016)) contain approximately double the number of reactions than the previous ones (iSB619 and iMH55) published in 2005. This exemplifies how models for the same organism have become larger in size with time, mainly due to a greater amount of genetic and metabolic data available for the organisms and the improvement of genome annotations. However, it is important to note that the accuracy and capability of a system's network cannot be solely determined by the number of reactions or metabolites that it contains, since a large proportion of reactions included in a network could still be unable to carry flux at steady state and metabolites could still be 'orphan' (involved in just a single reaction or only produced or consumed).

Some discrepancies were observed between model predictions and reported experimental results for iSB619 and iMH551. These were partly attributed to the lack of integration of regulatory data in the models, and the fact that, by the time they were constructed, a completed genome annotation for *S. aureus* N315 was not yet available. We identified a lack of consensus in the quality checks applied to these models, as well as in the way some of their main properties are described; this was especially noticeable in the earlier models. This has led to inconsistencies between publications and even data presented in ambiguous manners: e.g. *iSB619* was described as containing 640 metabolic reactions by Becker and Palsson (2005) and later described as consisting of 742 reactions by Bosi *et al.* (2016). These issues, together with the lack of standardised quality checks applied to the systems, are being tackled by current efforts in the field and are on their way to being resolved (Kumar *et al.* 2012; Ravikrishnan *et al.* 2015). The model constructed during this project was curated working in a best practice manner. This is reflected in the system's general properties, identified and described in Chapter 2, Table 2-1.

## 7.2 Results overview

In this project, a genome-scale metabolic model for *S. epidermidis* RP62A has been built, subjected to extensive manual curation, experimentally validated in house and against published biochemical data and, finally, successfully utilised to explore the metabolic mechanisms that enable this strain to produce energy, planktonic biomass and PIA (the main biofilm exopolysaccharide) in human joints:

In Chapter 2 the process of model construction and curation was described in detail. Subjecting the GSM to extensive quality checks ensured it was conserved for mass, energy and redox, and free from stoichiometric inconsistencies.

The focus of Chapter 3 was to perform further analysis of the system in order to define its ability to reproduce well-known physiological characteristics of the modelled-organism, such as its capacity to produce biomass, to utilise both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> as electron acceptors, and to produce ATP upon different conditions in a reasonable manner. While in Chapter 2 the model was shown to account for production of single biomass components, in Chapter 3 it was proven to be capable of doing so in the experimentally described proportions for staphylococci. The characterisation of the ETC performed in Chapter 3 showed that both O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> could be utilised as electron acceptors in responses that presented reasonable ATP/NADH and P/O ratios. Analysis of the system for ATP production proved its ability to mimic well-known metabolic strategies for staphylococci, such as: i) production of energy by metabolising Glc to Ac while repressing flux through reactions of the TCA cycle in the presence of both Glc and O<sub>2</sub> (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016; Halsey *et al.* 2017); ii) utilisation of Ac via the TCA cycle (prior conversion to AcCoA) in Glc-depleted media (Somerville *et al.* 2003; Sadykov *et al.* 2013; Somerville 2016); iii) fermentation of Glc in the absence of electron acceptors (Sivakanesan *et al.* 1980; Resch *et al.* 2005; Fuchs *et al.* 2007; Zhu *et al.* 2007) ; and iv) use of acetoin and butanediol for the synthesis of ATP (Yao *et al.* 2005; Cassat *et al.* 2006; Xiao *et al.* 2007; Zhu *et al.* 2007). These responses not only reproduced the

organism's behaviour observed *in vitro* but also provided a possible metabolic explanation for their implementation, since these pathways were chosen by the network when minimisation of total net flux through the system was selected as the objective function of the analyses. Having considered this parameter as a proxy for enzymatic investment, these results seemed to confirm protein economy as a plausible biological objective for cells living under these conditions. Therefore, using LP-based analysis results as a basis for defining regulatory patterns to be exerted by RP62A *in vitro* and *in vivo* appears plausible, as long as a realistic objective is considered during the analyses. It is, however, not yet possible to accurately define the biological objective/s followed by cells growing under different conditions, hence we can only hypothesise or speculate which these might be. This highlights the importance of validating the solutions obtained during model analysis by performing the necessary experiments to help us determine if the selected biological objectives represent a valid choice or not. Finally, analysing the model for production of planktonic biomass identified synthesis and excretion of autoinducer 2, a quorum sensing signalling molecule which causes cellular gene expression to respond to changes in cell-density (Miller *et al.* 2001; Zhu *et al.* 2003), as an inevitable consequence of synthesising menaquinones for cell growth (Chapter 3, Section 3.4.4). Perhaps, indicating that LP-based analysis could help establishing links between cellular metabolism and certain mechanisms for control of gene expression.

Once the quality of the model and its consistency with the physiological behaviour of the organism was established, further work was performed in order to ensure that the system accurately reflected the organism's ability to synthesise amino acids. Since amino acid utilisation is an important feature of biofilm metabolism (Zhu *et al.* 2007), this seemed a necessary step before applying model analysis to the study of biofilm formation. Hence Chapter 4 focused on the acquisition and analysis of experimental data for minimal-growth requirements, which was used for curation of reactions leading to the production of amino acids included in cell biomass. Briefly, the organism was cultured in a chemically defined medium from which a single amino acid was removed at a time and the effect of this removal on cell growth was compared with LP-based analysis results reproducing these growth conditions *in silico*. Extensive curation was performed based on these experimental results, together with the analysis of the genetic content of the organism. Although the sole contradiction found between the *in vitro* and the *in silico* results laid in the absence of growth without Pro observed *in vitro* but not *in silico*, further experimental work described in Chapter 5 demonstrated that RP62A is in fact non-auxotrophic for Pro, as the model analysis had accurately described. Some discrepancies observed between these experimental results and previously published experimental data for RP62A were explained as a consequence of changes in gene regulation during an adaptive response to a more nutrient-restricted environment, with similar effects previously reported in staphylococci (Gladstone 1937; Knight 1937; Emmett *et al.* 1975; Heinemann *et al.* 2005; Lee *et al.* 2009; Bosi *et al.* 2016). The effect of amino acid removal on biofilm formation was also measured experimentally and compared to results obtained by computing the total net flux through the system and Glc uptake associated to the corresponding *in silico* responses for growth

under these conditions. The changes observed in these parameters appeared to be insufficient to explain the *in vitro* effects observed on cell growth and the production of biofilm. There are several factors that are not taken into account in structural models (e.g. enzyme kinetics, expression of enzymatic genes and complex regulatory events such as those regulating biofilm formation), therefore, even though GSMs are useful tools to explore and predict flux patterns related to events or conditions of interest, the incorporation (in time) of kinetic parameters, transcriptomic and/or proteomic data together with improved genome annotations and better information on cellular transporters will increase the accuracy and the predictive power of the analyses of these models.

The main focus of Chapter 5 was to apply LP-based analysis to the study of N metabolism and amino acid utilisation for biomass production. The results obtained informed the experimental design used to produce further data for validation purposes. This work demonstrated how results derived from model analysis can lead to the formulation of hypotheses about the modelled-organism and help design the experiments necessary to test them. For example, *in silico* results obtained in Chapter 4 during the investigation of minimal requirements for growth suggested RP62A as being capable of biosynthesizing Pro, and *in silico* results from Chapter 5 identified which three amino acids contributed to the total N uptake for biomass production in the highest proportion (Ala, Arg and Glt). With this information, a new medium was designed, containing large amounts of these three amino acids as sole N sources. This finally allowed growth in the absence of Pro, which was also partly attributed to Arg and Glt, precursors for ornithine needed for the synthesis of Pro, being available at high concentrations in this new media formulation. Further hypotheses were also formulated and successfully tested, such as the lack of ability for this strain to grow in medium containing 7 amino acids defined by LP-based analysis as non-suitable N sources for growth. However, others were not reproducible *in vitro*, such as growth in the sole presence of either Ala, Arg, Glt or  $\text{NH}_4^+$  as single N sources. This was partly explained by the increase in the total net flux through the system associated with these *in silico* responses in comparison to that obtained in standard MM medium. However, these values were not much higher than that obtained in medium with Ala, Arg and Glt combined. Therefore, this explanation was again complemented with a possible effect of regulatory events on amino acid metabolism during adaptation to nutrient-restricted media. Finally, results derived from the investigation of Glc and amino acid utilisation for *in silico* biomass production suggested that amino acid catabolism is tightly coupled with ATP production even in the presence of Glc, which is again reinforced by results described in Chapter 6.

Analyses described in Chapter 6 identified routes for production of ATP, PIA and planktonic biomass in synovial fluid under a range of conditions that could potentially be encountered by cells growing in joints. Single solutions were calculated in order to understand responses occurring under the most extreme conditions, such as total availability or absence of  $\text{O}_2$  and  $\text{NO}_3^-$ . The model was scanned for responses to changes in  $\text{O}_2$  levels, allowing identification of reactions that are important for adaptation to the conditions found across the structure of biofilms growing on prosthetic joints (Wimpenny *et al.* 1983; Wimpenny *et al.* 2000). Finally, calculating the elementary modes through

these responsive reactions identified potential routes leading to production of energy, PIA1 and PIA2. In this way, the metabolic network was shown to re-arrange itself and modify the utilisation of Glc and amino acids in response to environmental changes, in order to fulfil these demands. This work identified reactions not only involved with the ETC but also with the metabolism of Glt, Ser and citrulline as important for adaptation to energy production in the array of conditions found in joints: while the utilisation of Glt for ATP synthesis had been described in Chapter 3 and Chapter 5, the use of citrulline seemed to be not only new in the context of this project but also not described before for ATP synthesis in staphylococci living in joints. Regarding PIA production, the set of reactions metabolising AcCoA and Glc all the way to PIA1 and PIA2 remained unaltered in all solutions and accurately reflected the mechanisms involved in PIA synthesis described in the literature (Heilmann *et al.* 1996; Gerke *et al.* 1998; Lee *et al.* 2016; Somerville 2016). However, reactions from the ETC, glycolysis and those involved in the utilisation of Ala, Ser, citrulline or Glt were shown to be combined in different ways in order to comply with the ATP and AcCoA demand of the process, depending on the changing environmental conditions. The constant demand of Glc identified across these solutions is in line with the work of several authors reporting high Glc concentrations to induce biofilm formation *in vitro* (Vuong *et al.* 2005; Sadykov *et al.* 2008; Sadykov *et al.* 2011; Somerville 2016). Identification of a wide range of elementary modes leading to synthesis of PIA1 and PIA2 suggested that the ability of cells to modify the rate of acetylated and deacetylated residues in PIA might be of importance in controlling the characteristics of the biofilm matrix, which could be advantageous in the colonisation of surfaces presenting different charges. A similar phenomenon has been described *in vitro*, with staphylococci modifying the relative proportion of teichoic acids and PIA in the matrix with the same purpose (Gross *et al.* 2001; Sadovskaya *et al.* 2004). LP-based analysis showed that PIA synthesis is optimally achieved with the lowest associated total net flux through the system by avoiding flux through reactions of the TCA cycle and utilising alternative routes for production of energy and AcCoA. This coincides with the fact that external factors such as high Glc concentrations, nutrient and iron deprivation and decreased O<sub>2</sub> levels have been related to both repression of the TCA cycle and drastic *in vitro* increases in PIA production (Vuong *et al.* 2005; Sadykov *et al.* 2008; Sadykov *et al.* 2011; Somerville 2016), which supports the validity of these *in silico* results. This suggests that utilising minimisation of the total net flux through the system as a proxy for reduction of the protein investment is a plausible way to represent the real biological objective followed by cells living in biofilms. This work also provided an explanation for the marked reduction of biofilm formation observed *in vitro* in cultures grown without Glt (Chapter 4), whose catabolism had been shown to be important for ATP synthesis and could be a suitable source of Suc for succination of PIA residues. Finally, the metabolism of Glc and several amino acids (Ala, Arg, Glt, Ser and Thr) was also identified as relevant to the production of planktonic biomass under these conditions, with their utilisation as ATP and AcCoA sources for anabolic processes already been described for staphylococci *in vitro* (Sivakanesan *et al.* 1980; Tynecka *et al.* 1999; Halsey *et al.* 2017).

Metabolic modelling provides the advantage of applying a systems biology approach to the study of the metabolism of an organism, breaking pre-imposed biochemical assumptions and the traditional division of metabolic pathways between anabolic or catabolic. Commonly, it has been assumed that catabolic reactions exist to produce energy for anabolic reactions. What this work has shown is that, if we do not build this assumption into our analyses, a different picture can emerge, where other alternative routes for the production of energy are prioritised. This was exemplified by the results obtained for ATP production when: i) the conditions encountered by cells during the exponential growth phase of staphylococci in the presence of Glc and O<sub>2</sub> were reproduced, resulting in Glc metabolism being diverted to Ac production rather than to synthesis of AcCoA to be fuelled to the TCA cycle (Chapter 3, Section 3.3.2), and ii) the system was analysed for PIA production, where degradation of amino acids and the deacetylation of PIA contributed to fulfil the energy requirement of the process, rather than glycolysis, the TCA cycle or the PPP (Chapter 6, Section 6.3.2). It is becoming clearer and clearer that this approach is needed in order to understand the functioning of cells as a whole, and to obtain answers to questions such as why certain genes and not others are essential for particular processes, and why, for example, flux through certain reactions is prioritised above flux through others that would have eventually led to higher energy yields.

### **7.3 Future work**

Since the LP-based analyses involving biomass production performed during this work relied on an averaged biomass composition for *S. aureus* with some minor modifications specific to *S. epidermidis*, obtaining a specific biomass composition for *S. epidermidis* RP62A would improve the accuracy of the results obtained. Some macromolecular components have already been measured for RP62A cells growing in MM minimal medium by other members of the team (Dr Noemi Tejera-Hernandez) and this work will be finalised in the near future. For the same reason, obtaining an accurate biomass composition for biofilms grown in synovial fluid or even in MM minimal medium would also be very useful. In addition, defining the specific growth rate for this strain, and even its ATP demand for growth rather than using estimated parameters based on data derived from other organisms would improve the quality of the results derived from model analysis. A possible way to achieve this would be to grow RP62A in a minimal medium with Glc as sole C source and under a range of Glc concentrations, identifying the minimum Glc media concentration that supports growth. Measuring the amount of Glc present in the medium before and after growth and the amount of biomass produced in that time period would allow us to determine the amount of Glc consumed per gDW per time unit. Having calculated the ATP/Glc rate for the *in silico* response for growth under similar conditions (Chapter 3, Section 3.3.2.1), we could then calculate the minimal ATP demand for growth. However, this would only be possible providing that: i) the strain could be grown in minimum media containing NH<sub>4</sub><sup>+</sup> as sole N source (which has not been achieved so far), and, ii) transcriptomic data could be obtained during these experiments in order to confirm that Glc is

metabolised following the same routes as described in the corresponding LP solution if the expression of genes encoding the enzymes involved in the *in silico* response is seen to be upregulated.

Some alternative objective functions to minimisation of total net flux through the system applied during LP-based analysis of GSMs are maximisation of biomass production or minimisation of the number of reactions carrying flux, with this last one being considered as an alternative proxy for reduction of the enzymatic cost. These could possibly constitute an equally adequate or even a better match for the real biological objective followed by the biological organism, depending on the environmental conditions considered. The analyses performed in Chapter 4 and Chapter 5 with a focus on studying the effects of amino acid deprivation on growth and biofilm formation could be repeated using these alternative objective functions. This would allow us to, for example, investigate if the *in silico* effects detected on the objective values of these new solutions are in higher agreement with the effects observed *in vitro*, and therefore, a different objective function should be considered.

Repeating the *in vitro* testing for growth on different N sources (Chapter 5) but excluding Glc in the media, could help reveal if C catabolite repression has a negative effect on the utilisation of amino acids for growth when single amino acids are provided as N sources, or if this prevents the synthesis of amino acids from  $\text{NH}_4^+$ . If that is the case, the organism should be able to synthesise amino acids more effectively in the absence of Glc, allowing concordance between *in silico* and *in vitro* results.

The importance of the metabolic routes defined in Chapter 6 for production of ATP, PIA and planktonic biomass under a range of environmental conditions could be confirmed experimentally by obtaining transcriptomic and proteomic data from cultures growing in MM medium (or even in artificial synovial fluid) under a range of  $\text{O}_2$  concentrations. Increasing the Glc concentration in the medium could be a plausible way to promote biofilm formation for the study of different growth phenotypes during these experiments (planktonic *vs* biofilm). In a similar fashion, obtaining transcriptomic and proteomic data on cells grown after several passages in MM minimal medium in anaerobiosis, with and without added citrulline, could show if genes/proteins involved in the uptake and catabolism of this compound are overexpressed, thus confirming or disproving the importance of citrulline utilisation in the absence of electron acceptors, as was suggested in this chapter.

Since the *in silico* results suggest that RP62A cells could adapt to grow in joints (Chapter 6), it would be reasonable to hypothesise that this organism could produce the enzymes needed for digestion of hyaluronic acid into its monomeric components (glucuronic acid and NAcGlc) as well as for direct phosphorylation of NAcGlc into NAcGlc-1P, to be utilised for the synthesis of UDP-NAcGlc and subsequently PIA. It is possible that these enzymes are missing from the PGDB for RP62A due to an incorrect or incomplete genome annotation. These hypotheses could be verified experimentally by feeding a bacterial culture with hyaluronic acid and analysing the medium composition before

and after growth, measuring the levels of unmodified hyaluronic acid and of its digestion products. In the same manner, if NAcGlc was provided with the medium instead, the ability of the organism to metabolise this compound could also be determined.

A next step in the study of the metabolism of cells living in biofilms would be to analyse the model for responses that satisfy both the energy demand for GAM and NGAM and the production of PIA combined. This would provide a more complete picture of the metabolic processes that could be taking place on these cells. Analysing the model for the identification of essential reactions for PIA production in synovial fluid would allow a selection of genes to be proposed as potential new drug targets, and could be achieved by identifying which reactions are part of the same reaction subset as the PIA1 and PIA2 exporter, and therefore, if blocked, would render the system unable to export these products. Another way to define essential reactions for PIA synthesis would be to perform single and double reaction knockout analysis, thus defining single reactions and reaction pairs which are essential for the process. Performing further model analysis simulating an increase in the PIA demand while maintaining constant the constraints for GAM and NGAM energy costs and planktonic biomass production could be another approach for determining reactions involved in the adaptation of cells to grow in biofilms.

Since it is known that cells encounter a gradient of nutrients and electron acceptors throughout the biofilm, it is highly likely that those growing or surviving on the lower layers of a biofilm do so by metabolising by-products excreted by cells on the upper layers (e.g. Ac, Form or Suc), rather than by catabolising compounds present in the synovial fluid. Hence it would be interesting to perform similar analyses to those described in Chapter 3, Section 3.4.3.2 for the study of the utilisation of metabolic by-products for ATP production, investigating synthesis of PIA from these by-products, as well as the potential cross-feeding between planktonic and biofilm cells.

## **7.4 Conclusions**

Since mathematical modelling of metabolism led to publishing the first genome-scale metabolic model of a living-organism in 2000 (Schilling *et al.* 2000), modelling of large-scale metabolic networks has advanced at a slow but steady pace. Current efforts in the field are focused on the development of new techniques for the analysis of metabolic models, automating and increasing the scalability of the construction and curation steps, the integration of large -omic data sets and the construction of genome-scale kinetic models. In reality, building a highly accurate GSM remains a fairly skilled, difficult and time-consuming task. However, these models are powerful tools with many possible applications, that once constructed and curated, can be used to identify and interpret metabolic processes, perform ‘what-if’ analyses where multiple scenarios can be quickly reproduced, and guide the experimental work needed to corroborate these findings, thus reducing substantially the time and resources invested at this end of the process.

Despite the increasing pathogenic importance of *S. epidermidis* (Becker *et al.* 2014; Uribe-Alvarez *et al.* 2016; Pedroza-Davila *et al.* 2020), little is still known about the metabolic pathways that allow it to survive and produce biofilms. This project has demonstrated that genome-scale metabolic modelling can be successfully applied to this aim: analysis of the GSM for RP62A built here has not only been able to reproduce well-known aspects of the organism's physiology but has also highlighted important pathways for production of energy, planktonic biomass and biofilm components in joints. While the involvement of some reactions was expected, as those responsible for metabolising AcCoA and Glc-6P all the way down to PIA, the catabolism of citrulline in the absence of electron acceptors as a way to satisfy the ATP demand of the process is new in this context. The results derived from this work have substantially contributed to our overall understanding on how cells adapt to changing conditions across the biofilm structure, re-arranging their network and varying the utilisation of different substrates. There is, however, much else that could be done such as: i) analysing the model for PIA together with other biofilm components, which will complement the current results and could establish connections between areas of the network that might be impossible to define otherwise; ii) defining reactions essential for these processes, which could lead to the identification of novel antibiofilm drug targets or candidate genes to be knocked-out in order to obtain virulence-attenuated strains for the development of vaccines against NAS infections; iii) identifying gene-knockouts that could increase production of biofilm-matrix components in order to exploit this organism for the production of commercial products such as antibiofilm vaccines (Gil *et al.* 2014; Somerville 2016); or even iv) constructing pangenome GSMs for NAS isolates included in the same RC group (Chapter 0, Section 1.3.2) and comparing them in order to study the metabolic diversity of NAS species and as a way to define if there are relevant metabolic features that differentiate these clusters, helping to improve or validate this typing scheme.

This PhD project has not only successfully applied metabolic modelling to the study of the most important NAS species involved in PJI, but has also allowed the development of strong collaborative links with the Cell System Modelling Group at Oxford Brookes University while strongly contributing to bringing further expertise in the field to the Norwich Research Park, consequently expanding the application of these techniques to other research projects.

## 8 Bibliography

- Agarwal, A. and A. Jain (2013). "Glucose & sodium chloride induced biofilm production & ica operon in clinical isolates of staphylococci." Indian J Med Res **138**(2): 262-266.
- Aggarwal, V. K., H. Bakhshi, N. U. Ecker, J. Parvizi, T. Gehrke and D. Kendoff (2014). "Organism profile in periprosthetic joint infection: pathogens differ at two arthroplasty infection referral centers in Europe and in the United States." The journal of knee surgery **27**(05): 399-406.
- Allignet, J., J.-O. Galdbart, A. Morvan, K. G. Dyke, P. Vaudaux, S. Aubert, N. Desplaces and N. El Solh (1999). "Tracking adhesion factors in *Staphylococcus caprae* strains responsible for human bone infections following implantation of orthopaedic material." Microbiology **145**(8): 2033-2042.
- Amon, J., F. Titgemeyer and A. Burkovski (2010). "Common patterns - unique features: nitrogen metabolism and regulation in Gram-positive bacteria." FEMS Microbiol Rev **34**(4): 588-605.
- Anraku, Y. (1988). "Bacterial electron transport chains." Annu Rev Biochem **57**: 101-132.
- Arciola, C. R., Y. An, D. Campoccia, M. E. Donati and L. Montanaro (2005). "Etiology of implant orthopedic infections: a survey on 1027 clinical isolates." The International journal of artificial organs **28**(11): 1091-1100.
- Arciola, C. R., D. Campoccia, P. Speziale, L. Montanaro and J. W. Costerton (2012). "Biofilm formation in *Staphylococcus* implant infections. A review of molecular mechanisms and implications for biofilm-resistant materials." Biomaterials **33**(26): 5967-5982.
- Arciola, C. R., L. Visai, F. Testoni, S. Arciola, D. Campoccia, P. Speziale and L. Montanaro (2011). "Concise survey of *Staphylococcus aureus* virulence factors that promote adhesion and damage to peri-implant tissues." The International journal of artificial organs **34**(9): 771-780.
- Argemi, X., P. Riegel, T. Lavigne, N. Lefebvre, N. Grandpré, Y. Hansmann, B. Jaulhac, G. Prévost and F. Schramm (2015). "Implementation of matrix-assisted laser desorption ionization–time of flight mass spectrometry in routine clinical laboratories improves identification of coagulase-negative staphylococci and reveals the pathogenic role of *Staphylococcus lugdunensis*." Journal of clinical microbiology **53**(7): 2030-2036.
- Atkins, B. L., N. Athanasou, J. J. Deeks, D. W. Crook, H. Simpson, T. E. Peto, P. McLardy-Smith, A. R. Berendt and O. C. S. Group (1998). "Prospective evaluation of criteria for microbiological diagnosis of prosthetic-joint infection at revision arthroplasty." Journal of clinical microbiology **36**(10): 2932-2939.
- Baldan, R., F. Testa, N. I. Lorè, A. Bragonzi, P. Cichero, C. Ossi, A. Biancardi, P. Nizzero, M. Moro and D. M. Cirillo (2012). "Factors Contributing to Epidemic MRSA Clones Replacement in a Hospital Setting." PLOS ONE **7**(8): e43153.
- Baldassarri, L., G. Donelli, A. Gelosia, M. C. Voglino, A. W. Simpson and G. D. Christensen (1996). "Purification and characterization of the staphylococcal slime-associated antigen and its occurrence among *Staphylococcus epidermis* clinical isolates." Infect Immun **64**(8): 3410-3415.
- Basan, M., S. Hui, H. Okano, Z. Zhang, Y. Shen, J. R. Williamson and T. Hwa (2015). "Overflow metabolism in *Escherichia coli* results from efficient proteome allocation." Nature **528**(7580): 99-104.
- Bauchop, T. and S. R. Elsdén (1960). "The growth of micro-organisms in relation to their energy supply." J Gen Microbiol **23**: 457-469.

- Becker, K., D. Harmsen, A. Mellmann, C. Meier, P. Schumann, G. Peters and C. Von Eiff (2004). "Development and evaluation of a quality-controlled ribosomal sequence database for 16S ribosomal DNA-based identification of Staphylococcus species." Journal of clinical microbiology **42**(11): 4988-4995.
- Becker, K., C. Heilmann and G. Peters (2014). "Coagulase-negative staphylococci." Clin Microbiol Rev **27**(4): 870-926.
- Becker, S. A. and B. O. Palsson (2005). "Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation." BMC microbiology [electronic resource] **5**.
- Becker, S. A. and B. Ø. Palsson (2005). "Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation." BMC Microbiology **5**(1): 8.
- Beenken, K. E., P. M. Dunman, F. McAleese, D. Macapagal, E. Murphy, S. J. Projan, J. S. Blevins and M. S. Smeltzer (2004). "Global gene expression in Staphylococcus aureus biofilms." J Bacteriol **186**(14): 4665-4684.
- Belmatoug, N., A.-C. Crémieux, R. Bleton, A. Volk, A. Saleh-Mghir, M. Grossin, L. Garry and C. Carbon (1996). "A new model of experimental prosthetic joint infection due to methicillin-resistant Staphylococcus aureus: a microbiologic, histopathologic, and magnetic resonance imaging characterization." Journal of Infectious Diseases **174**(2): 414-417.
- Bémer, P., J. Léger, D. Tandé, C. Plouzeau, A. S. Valentin, A. Jolivet-Gougeon, C. Lemarié, M. Kempf, G. Héry-Arnaud and L. Bret (2016). "How many samples and how many culture media to diagnose a prosthetic joint infection: a clinical and microbiological prospective multicenter study." Journal of clinical microbiology **54**(2): 385-391.
- Benito, N., M. Franco, A. Ribera, A. Soriano, D. Rodriguez-Pardo, L. Sorlí, G. Fresco, M. Fernández-Sampedro, M. D. del Toro and L. Guío (2016). "Time trends in the aetiology of prosthetic joint infections: a multicentre cohort study." Clinical Microbiology and Infection **22**(8): 732. e731-732. e738.
- Blumenthal, H. J. (1972). "Glucose catabolism in staphylococci." The staphylococci. Wiley-Interscience, New York, NY: 111-135.
- Bosi, E., J. M. Monk, R. K. Aziz, M. Fondi, V. Nizet and B. O. Palsson (2016). "Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity." Proc Natl Acad Sci U S A **113**(26): E3801-3809.
- Boutte, C. C. and S. Crosson (2013). "Bacterial lifestyle shapes stringent response activation." Trends Microbiol **21**(4): 174-180.
- Bradford, N. M. and J. D. McGivan (1980). "Evidence for the existence of an ornithine/citrulline antiporter in rat liver mitochondria." FEBS Lett **113**(2): 294-298.
- Brauner, A., O. Fridman, O. Gefen and N. Q. Balaban (2016). "Distinguishing between resistance, tolerance and persistence to antibiotic treatment." Nat Rev Microbiol **14**(5): 320-330.
- Bückle, A., M. Kranz, H. Schmidt and A. Weiss (2017). "Genetic diversity and population structure of food-borne Staphylococcus carnosus strains." Systematic and Applied Microbiology **40**(1): 34-41.
- Burke, K. A. and J. Lascelles (1975). "Nitrate reductase system in Staphylococcus aureus wild type and mutants." J Bacteriol **123**(1): 308-316.
- Burton, E., P. V. Gawande, N. Yakandawala, K. LoVetri, G. G. Zhanel, T. Romeo, A. D. Friesen and S. Madhyastha (2006). "Antibiofilm activity of GlmU enzyme inhibitors against catheter-associated uropathogens." Antimicrob Agents Chemother **50**(5): 1835-1840.

- Büttner, H., D. Mack and H. Rohde (2015). "Structural basis of Staphylococcus epidermidis biofilm formation: mechanisms and molecular interactions." Frontiers in Cellular and Infection Microbiology **5**: 14.
- Campoccia, D., L. Montanaro, L. Visai, T. Corazzari, C. Poggio, F. Pegreff, A. Maso, V. Pirini, S. Ravaioli and I. Cangini (2010). "Characterization of 26 Staphylococcus warneri isolates from orthopedic infections." International Journal of Artificial Organs **33**(9): 575-581.
- Canepa, A., J. C. Filho, A. Gutierrez, A. Carrea, A. M. Forsberg, E. Nilsson, E. Verrina, F. Perfumo and J. Bergstrom (2002). "Free amino acids in plasma, red blood cells, polymorphonuclear leukocytes, and muscle in normal and uraemic children." Nephrol Dial Transplant **17**(3): 413-421.
- Carbonnelle, E., J.-L. Beretti, S. Cottyn, G. Quesne, P. Berche, X. Nassif and A. Ferroni (2007). "Rapid identification of Staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry." Journal of clinical microbiology **45**(7): 2156-2161.
- Carlson, R., A. Beck, P. Phalak, M. Fields, T. Gedeon, L. Hanley, W. Harcombe, M. Henson and J. Heys (2018). "Competitive resource allocation to metabolic pathways contributes to overflow metabolisms and emergent properties in cross-feeding microbial consortia." Biochemical Society Transactions **46**: BST20170242.
- Carver, T., S. R. Harris, M. Berriman, J. Parkhill and J. A. McQuillan (2011). "Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data." Bioinformatics **28**(4): 464-469.
- Carver, T. J., K. M. Rutherford, M. Berriman, M.-A. Rajandream, B. G. Barrell and J. Parkhill (2005). "ACT: the Artemis comparison tool." Bioinformatics **21**(16): 3422-3423.
- Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp (2014). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases." Nucleic Acids Res **42**.
- Caspi, R., T. Altman, K. Dreher, C. A. Fulcher, P. Subhraveti, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, A. Pujar, A. G. Shearer, M. Travers, D. Weerasinghe, P. Zhang and P. D. Karp (2012). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases." Nucleic acids research **40**.
- Caspi, R., R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver and P. D. Karp (2016). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases." Nucleic Acids Res **44**(D1): D471-480.
- Cassat, J., P. M. Dunman, E. Murphy, S. J. Projan, K. E. Beenken, K. J. Palm, S.-J. Yang, K. C. Rice, K. W. Bayles and M. S. Smeltzer (2006). "Transcriptional profiling of a Staphylococcus aureus clinical isolate and its isogenic agr and sarA mutants reveals global differences in comparison to the laboratory strain RN6390." Microbiology **152**(10): 3075-3090.
- Christensen, G. D., W. Simpson, J. Younger, L. Baddour, F. Barrett, D. Melton and E. Beachey (1985). "Adherence of coagulase-negative staphylococci to plastic tissue culture plates: a quantitative model for the adherence of staphylococci to medical devices." Journal of clinical microbiology **22**(6): 996-1006.
- Chu, Y. C., C. Z. Chen, C. H. Lee, C. W. Chen, H. Y. Chang and T. R. Hsiue (2003). "Prediction of arterial blood gas values from venous blood gas values in patients with acute respiratory failure receiving mechanical ventilation." J Formos Med Assoc **102**(8): 539-543.

- Conlon, K. M., H. Humphreys and J. P. O'Gara (2002). "icaR encodes a transcriptional repressor involved in environmental regulation of ica operon expression and biofilm formation in *Staphylococcus epidermidis*." J Bacteriol **184**(16): 4400-4408.
- Cowan, S., C. Shaw and R. Williams (1954). "Type strain for *Staphylococcus aureus* Rosenbach." Microbiology **10**(1): 174-176.
- Cue, D., M. G. Lei and C. Y. Lee (2012). "Genetic regulation of the intercellular adhesion locus in staphylococci." Front Cell Infect Microbiol **2**: 38.
- Curnow, A. W., D. L. Tumbula, J. T. Pelaschier, B. Min and D. Soll (1998). "Glutamyl-tRNA(Gln) amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis." Proc Natl Acad Sci U S A **95**(22): 12838-12843.
- De Kievit, T. R., R. Gillis, S. Marx, C. Brown and B. H. Iglewski (2001). "Quorum-sensing genes in *Pseudomonas aeruginosa* biofilms: their role and expression patterns." Appl Environ Microbiol **67**(4): 1865-1873.
- de Langue Française, S. P. I. (2010). "Recommendations for bone and joint prosthetic device infections in clinical practice (prosthesis, implants, osteosynthesis)." Médecine et Maladies Infectieuses **40**(4): 185-211.
- Dengler Haunreiter, V., M. Boumasmoud, N. Haffner, D. Wipfli, N. Leimer, C. Rachmuhl, D. Kuhnert, Y. Achermann, R. Zbinden, S. Benussi, C. Vulin and A. S. Zinkernagel (2019). "In-host evolution of *Staphylococcus epidermidis* in a pacemaker-associated endocarditis resulting in increased antibiotic tolerance." Nat Commun **10**(1): 1149.
- Deutscher, J. (2008). "The mechanisms of carbon catabolite repression in bacteria." Curr Opin Microbiol **11**(2): 87-93.
- Dufour, P., S. Jarraud, F. Vandenesch, T. Greenland, R. P. Novick, M. Bes, J. Etienne and G. Lina (2002). "High genetic variability of the agr locus in *Staphylococcus* species." Journal of bacteriology **184**(4): 1180-1186.
- Durot, M., P. Y. Bourguignon and V. Schachter (2009). "Genome-scale models of bacterial metabolism: reconstruction and applications." FEMS Microbiol Rev **33**(1): 164-190.
- Edwards, J. S., M. Covert and B. Palsson (2002). "Metabolic modelling of microbes: the flux-balance approach." Environ Microbiol **4**.
- Edwards, J. S. and B. O. Palsson (1999). "Systems properties of the *Haemophilus influenzae* Rd metabolic genotype." J Biol Chem **274**(25): 17410-17416.
- El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto and R. D. Finn (2018). "The Pfam protein families database in 2019." Nucleic Acids Research **47**(D1): D427-D432.
- El-Mansi, M. (2004). "Flux to acetate and lactate excretions in industrial fermentations: physiological and biochemical implications." J Ind Microbiol Biotechnol **31**(7): 295-300.
- Emmett, M. and W. E. Kloos (1975). "Amino acid requirements of staphylococci isolated from human skin." Canadian journal of microbiology **21**(5): 729-733.
- Fairbrother, R. (1940). "Coagulase production as a criterion for the classification of the staphylococci." The Journal of Pathology and Bacteriology **50**(1): 83-88.
- Feist, A. M., C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. Palsson (2007). "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol **3**: 121.
- Feist, A. M., M. J. Herrgard, I. Thiele, J. L. Reed and B. O. Palsson (2009). "Reconstruction of biochemical networks in microorganisms." Nat Rev Micro **7**(2): 129-143.
- Fell, D. A. and J. R. Small (1986). "Fat synthesis in adipose tissue. An examination of stoichiometric constraints." Biochemical Journal **238**(3): 781-786.

- Fernanda Cristina Possamai Rossatto, J. B. P., Géssica Aracéli Costa and Ana Paula Guedes Frazzon (2017). "In vitro biofilm formation ability of staphylococci under different growth conditions." International Journal of Applied Microbiology and Biotechnology Research **5**(2): 12-19.
- Ferrier, D. R. (2014). Biochemistry. Philadelphia, Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Flamholz, A., E. Noor, A. Bar-Even and R. Milo (2012). "eQuilibrator--the biochemical thermodynamics calculator." Nucleic Acids Res **40**(Database issue): D770-775.
- Fluckiger, U., C. Wolz and A. L. Cheung (1998). "Characterization of a sar Homolog of Staphylococcus epidermidis." Infection and Immunity **66**(6): 2871-2878.
- Fong, S. S., A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas and B. O. Palsson (2005). "In silico design and adaptive evolution of Escherichia coli for production of lactic acid." Biotechnol Bioeng **91**(5): 643-648.
- Frame, E. G. (1958). "The Levels of Individual Free Amino Acids in the Plasma of Normal Man at Various Intervals After a High-Protein Meal." J Clin Invest **37**(12): 1710-1723.
- Fu, J., M. Ni, W. Chai, X. Li, L. Hao and J. Chen (2019). "Synovial Fluid Viscosity Test is Promising for the Diagnosis of Periprosthetic Joint Infection." J Arthroplasty **34**(6): 1197-1200.
- Fuchs, S., J. Pane-Farre, C. Kohler, M. Hecker and S. Engelmann (2007). "Anaerobic gene expression in Staphylococcus aureus." J Bacteriol **189**(11): 4275-4289.
- Gale, L. R. (2007). Biotribological assessment for artificial synovial joints : the role of boundary lubrication. Doctor of Philosophy PhD, Queensland University of Technology, Brisbane.
- Garrett, R. H. G., C.M. (2010). Biochemistry, Brooks/Cole.
- Geiger, T. and C. Wolz (2014). "Intersection of the stringent response and the CodY regulon in low GC Gram-positive bacteria." Int J Med Microbiol **304**(2): 150-155.
- Gerke, C., A. Kraft, R. Sussmuth, O. Schweitzer and F. Gotz (1998). "Characterization of the N-acetylglucosaminyltransferase activity involved in the biosynthesis of the Staphylococcus epidermidis polysaccharide intercellular adhesin." J Biol Chem **273**(29): 18586-18593.
- Gevorgyan, A., M. G. Poolman and D. A. Fell (2008). "Detection of stoichiometric inconsistencies in biomolecular models." Bioinformatics **24**(19): 2245-2251.
- Ghasemi, A., S. Zahediasl and F. Azizi (2010). "Reference values for serum nitric oxide metabolites in an adult population." Clin Biochem **43**(1-2): 89-94.
- Ghebremedhin, B., F. Layer, W. König and B. König (2008). "Genetic classification and distinguishing of Staphylococcus species based on different partial gap, 16S rRNA, hsp60, rpoB, sodA, and tuf gene sequences." Journal of clinical microbiology **46**(3): 1019-1025.
- Giacometti, A., O. Cirioni, Y. Gov, R. Ghiselli, M. S. Del Prete, F. Mocchegiani, V. Saba, F. Orlando, G. Scalise, N. Balaban and G. Dell'Acqua (2003). "RNA III inhibiting peptide inhibits in vivo biofilm formation by drug-resistant Staphylococcus aureus." Antimicrob Agents Chemother **47**(6): 1979-1983.
- Gil, C., C. Solano, S. Burgui, C. Latasa, B. García, A. Toledo-Arana, I. Lasa and J. Valle (2014). "Biofilm matrix exoproteins induce a protective immune response against Staphylococcus aureus biofilm infection." Infect Immun **82**(3): 1017-1029.
- Gilchrist, M., A. C. Shore and N. Benjamin (2010). "Inorganic nitrate and nitrite and control of blood pressure." Cardiovascular Research **89**(3): 492-498.
- Gill, S. R., D. E. Fouts, G. L. Archer, E. F. Mongodin, R. T. Deboy, J. Ravel, I. T. Paulsen, J. F. Kolonay, L. Brinkac, M. Beanan, R. J. Dodson, S. C. Daugherty, R. Madupu, S. V. Angiuoli, A. S. Durkin, D. H. Haft, J. Vamathevan, H. Khouri, T. Utterback, C. Lee, G. Dimitrov, L. Jiang, H.

- Qin, J. Weidman, K. Tran, K. Kang, I. R. Hance, K. E. Nelson and C. M. Fraser (2005). "Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain." *J Bacteriol* **187**(7): 2426-2438.
- Gladstone, G. P. (1937). "The Nutrition of *Staphylococcus aureus*; Nitrogen Requirements." *British Journal of Experimental Pathology* **18**(4): 322-333.
- Gouy, M., S. Guindon and O. Gascuel (2009). "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." *Molecular Biology and Evolution* **27**(2): 221-224.
- Graham, J. E. and B. J. Wilkinson (1992). "Staphylococcus aureus osmoregulation: roles for choline, glycine betaine, proline, and taurine." *J Bacteriol* **174**(8): 2711-2716.
- Gross, M., S. E. Cramton, F. Gotz and A. Peschel (2001). "Key role of teichoic acid net charge in *Staphylococcus aureus* colonization of artificial surfaces." *Infect Immun* **69**(5): 3423-3426.
- H Kenneth Walker, M., W Dallas Hall, MD, and J Willis Hurst, MD. (1990). *Clinical Methods*. Boston, Butterworths.
- Halsey, C. R., S. Lei, J. K. Wax, M. K. Lehman, A. S. Nuxoll, L. Steinke, M. Sadykov, R. Powers and P. D. Fey (2017). "Amino Acid Catabolism in *Staphylococcus aureus* and the Function of Carbon Catabolite Repression." *mBio* **8**(1): e01434-01416.
- Halsey, C. R., S. Lei, J. K. Wax, M. K. Lehman, A. S. Nuxoll, L. Steinke, M. Sadykov, R. Powers and P. D. Fey (2017). "Amino Acid Catabolism in *Staphylococcus aureus* and the Function of Carbon Catabolite Repression." *mBio* **8**(1).
- Hartman, H. B., D. A. Fell, S. Rossell, P. R. Jensen, M. J. Woodward, L. Thorndahl, L. Jelsbak, J. E. Olsen, A. Raghunathan, S. Daepler and M. G. Poolman (2014). "Identification of potential drug targets in *Salmonella enterica* sv. Typhimurium using metabolic modelling and experimental validation." *Microbiology* **160**(Pt 6): 1252-1266.
- Heilmann, C., O. Schweitzer, C. Gerke, N. Vanittanakom, D. Mack and F. Gotz (1996). "Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*." *Mol Microbiol* **20**(5): 1083-1091.
- Heinemann, M., A. Kummel, R. Ruinatscha and S. Panke (2005). "In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network." *Biotechnol Bioeng* **92**(7): 850-864.
- Heinrich, R. and S. Schuster (1996). *The regulation of cellular systems*. New York, Hall.
- Hofmeyr, J. (1986). "Steady-state modelling of metabolic pathways: A guide for the prospective simulator." *Bioinformatics* **2**(1): 5-11.
- Holzhutter, H. G. (2006). "The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies." *Biosystems* **83**(2-3): 98-107.
- Hug, L. A., B. J. Baker, K. Anantharaman, C. T. Brown, A. J. Probst, C. J. Castelle, C. N. Butterfield, A. W. Hermsdorf, Y. Amano and K. Ise (2016). "A new view of the tree of life." *Nature Microbiology* **1**: 16048.
- Hussain, M., J. Hastings and P. White (1991). "A chemically defined medium for slime production by coagulase-negative staphylococci." *Journal of medical microbiology* **34**(3): 143-147.
- Ikuo, M., G. Nagano, Y. Saito, H. Mao, K. Sekimizu and C. Kaito (2014). "Inhibition of exotoxin production by mobile genetic element SCCmec-encoded psm-mec RNA is conserved in staphylococcal species." *PLoS One* **9**(6): e100260.
- Ito, T. and S. Yokoyama (2010). "Two enzymes bound to one transfer RNA assume alternative conformations for consecutive reactions." *Nature* **467**(7315): 612-616.

- Ives, P. R. and M. E. Bushell (1997). "Manipulation of the physiology of clavulanic acid production in *Streptomyces clavuligerus*." Microbiology **143 ( Pt 11)**: 3573-3579.
- Jamshidi, N. and B. Ø. Palsson (2007). "Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ 661 and proposing alternative drug targets." BMC Systems Biology **1(1)**: 26.
- Jarraud, S., C. Mougél, J. Thioulouse, G. Lina, H. Meugnier, F. Forey, X. Nesme, J. Etienne and F. Vandenesch (2002). "Relationships between *Staphylococcus aureus* genetic background, virulence factors, agr groups (alleles), and human disease." Infection and immunity **70(2)**: 631-641.
- Jeske, L., S. Placzek, I. Schomburg, A. Chang and D. Schomburg (2018). "BRENDA in 2019: a European ELIXIR core data resource." Nucleic Acids Research **47(D1)**: D542-D549.
- Kabimoldayev, I., A. D. Nguyen, L. Yang, S. Park, E. Y. Lee and D. Kim (2018). "Basics of genome-scale metabolic modeling and applications on C1-utilization." FEMS Microbiol Lett **365(20)**.
- Kallala, R. F., I. S. Vanhegan, M. S. Ibrahim, S. Sarmah and F. S. Haddad (2015). "Financial analysis of revision knee surgery based on NHS tariffs and hospital costs: does it pay to provide a revision service?" Bone Joint J **97-b(2)**: 197-201.
- Kanehisa, M. and S. Goto (2000). "KEGG: Kyoto encyclopedia of genes and genomes." Nucleic acids research **28**.
- Kanehisa, M., S. Goto, S. Kawashima and A. Nakaya (2002). "The KEGG databases at GenomeNet." Nucleic Acids Res **30**.
- Karp, P. D., R. Billington, R. Caspi, C. A. Fulcher, M. Latendresse, A. Kothari, I. M. Keseler, M. Krummenacker, P. E. Midford, Q. Ong, W. K. Ong, S. M. Paley and P. Subhraveti (2017). "The BioCyc collection of microbial genomes and metabolic pathways." Briefings in Bioinformatics **20(4)**: 1085-1093.
- Karp, P. D., M. Latendresse and R. Caspi (2011). "The pathway tools pathway prediction algorithm." Stand Genomic Sci **5(3)**: 424-429.
- Karp, P. D., M. Latendresse, S. M. Paley, M. Krummenacker, Q. D. Ong, R. Billington, A. Kothari, D. Weaver, T. Lee, P. Subhraveti, A. Spaulding, C. Fulcher, I. M. Keseler and R. Caspi (2016). "Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology." Brief Bioinform **17(5)**: 877-890.
- Karp, P. D., S. Paley and P. Romero (2002). "The Pathway Tools software." Bioinformatics **18(suppl\_1)**: S225-S232.
- Kizil, M. (2010). "A holistic approach for the problem of storage space allocation in cellular manufacturing." The International Journal of Advanced Manufacturing Technology **48(9)**: 1031-1043.
- Knight, B. C. J. G. (1937). "The nutrition of *Staphylococcus aureus*; nicotinic acid and vitamin B1." Biochemical Journal **31(5)**: 731.
- Kumar, A., P. F. Suthers and C. D. Maranas (2012). "MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases." BMC Bioinformatics **13**: 6.
- Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi and K. Hiramatsu (2001). "Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*." Lancet **357**.
- Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K.-i. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A.

- Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R.-i. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi and K. Hiramatsu (2001). "Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*." The Lancet **357**(9264): 1225-1240.
- Lange, J., A. Troelsen, R. W. Thomsen and K. Søballe (2012). "Chronic infections in hip arthroplasties: comparing risk of reinfection following one-stage and two-stage revision: a systematic review and meta-analysis." Clin Epidemiol **4**(4): 57-73.
- Le, K. Y., M. D. Park and M. Otto (2018). "Immune Evasion Mechanisms of *Staphylococcus epidermidis* Biofilm Infection." Front Microbiol **9**: 359.
- Lee, D.-S., H. Burd, J. Liu, E. Almaas, O. Wiest, A.-L. Barabási, Z. N. Oltvai and V. Kapatral (2009). "Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets." Journal of bacteriology **191**(12): 4015-4024.
- Lee, M. J., A. M. Geller, N. C. Bamford, H. Liu, F. N. Gravelat, B. D. Snarr, F. Le Mauff, J. Chabot, B. Ralph, H. Ostapska, M. Lehoux, R. P. Cerone, S. D. Baptista, E. Vinogradov, J. E. Stajich, S. G. Filler, P. L. Howell and D. C. Sheppard (2016). "Deacetylation of Fungal Exopolysaccharide Mediates Adhesion and Biofilm Formation." mBio **7**(2): e00252-00216.
- Li, C., F. Sun, H. Cho, V. Yelavarthi, C. Sohn, C. He, O. Schneewind and T. Bae (2010). "CcpA Mediates Proline Auxotrophy and Is Required for *Staphylococcus aureus* Pathogenesis." Journal of Bacteriology **192**(15): 3883-3892.
- Lim, Y., M. Jana, T. T. Luong and C. Y. Lee (2004). "Control of Glucose- and NaCl-Induced Biofilm Formation by *rbf* in *Staphylococcus aureus*." Journal of Bacteriology **186**(3): 722-729.
- Lin, S., R. E. Hanson and J. E. Cronan (2010). "Biotin synthesis begins by hijacking the fatty acid synthetic pathway." Nat Chem Biol **6**(9): 682-688.
- Lister, J. L. and A. R. Horswill (2014). "Staphylococcus aureus biofilms: recent developments in biofilm dispersal." Front Cell Infect Microbiol **4**.
- Lutz, M. (2001). Programming Python, O'Reilly & Associates.
- M. Lutz, D. A. (1999). Learning Python, O'Reilly & Associates.
- Mack, D., W. Fischer, A. Krokotsch, K. Leopold, R. Hartmann, H. Egge and R. Laufs (1996). "The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis." Journal of Bacteriology **178**(1): 175-183.
- Mack, D., N. Siemssen and R. Laufs (1992). "Parallel induction by glucose of adherence and a polysaccharide antigen specific for plastic-adherent *Staphylococcus epidermidis*: evidence for functional relation to intercellular adhesion." Infect Immun **60**(5): 2048-2057.
- Malatesha, G., N. K. Singh, A. Bharija, B. Rehani and A. Goel (2007). "Comparison of arterial and venous pH, bicarbonate, PCO<sub>2</sub> and PO<sub>2</sub> in initial emergency department assessment." Emerg Med J **24**(8): 569-571.
- McNamara, P. J. and R. A. Proctor (2000). "Staphylococcus aureus small colony variants, electron transport and persistent infections." Int J Antimicrob Agents **14**(2): 117-122.
- Miller, M. B. and B. L. Bassler (2001). "Quorum sensing in bacteria." Annu Rev Microbiol **55**: 165-199.
- Mincheol Kim, J. C. (2014). Chapter 4 - 16S rRNA Gene-Based Identification of Bacteria and Archaea using the EzTaxon Server. Methods in Microbiology. I. S. Michael Goodfellow, Jongsik Chun,, Academic Press. **41**: 61-74.

- Miragaia, M., J. Thomas, I. Couto, M. Enright and H. De Lencastre (2007). "Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data." Journal of bacteriology **189**(6): 2540-2552.
- Molenaar, D., R. van Berlo, D. de Ridder and B. Teusink (2009). "Shifts in growth strategies reflect tradeoffs in cellular economics." Mol Syst Biol **5**: 323.
- Montanaro, L., D. Campoccia and C. R. Arciola (2007). "Advancements in molecular epidemiology of implant infections and future perspectives." Biomaterials **28**(34): 5155-5168.
- Montanaro, L., P. Speziale, D. Campoccia, S. Ravaioli, I. Cangini, G. Pietrocola, S. Giannini and C. R. Arciola (2011). "Scenery of *Staphylococcus* implant infections in orthopedics." Future Microbiology **6**(11): 1329-1349.
- Montanaro, L., F. Testoni, A. Poggi, L. Visai, P. Speziale and C. R. Arciola (2011). "Emerging pathogenetic mechanisms of the implant-related osteomyelitis by *Staphylococcus aureus*." The International journal of artificial organs **34**(9): 781-788.
- Morgenstern, M., C. Erichsen, S. Hackl, J. Mily, M. Militz, J. Friederichs, S. Hungerer, V. Bühren, T. F. Moriarty and V. Post (2016). "Antibiotic resistance of commensal *Staphylococcus aureus* and coagulase-negative staphylococci in an international cohort of surgeons: a prospective point-prevalence study." PloS one **11**(2): e0148437.
- Nahaie, M. R., M. Goodfellow, D. E. Minnikin and V. Hajek (1984). "Polar lipid and isoprenoid quinone composition in the classification of *Staphylococcus*." J Gen Microbiol **130**(9): 2427-2437.
- National Joint Registry for England, W. a. N. I. (2016). "Knees - All Procedures - Activity." from <http://www.njrreports.org.uk/knees-reports>.
- Nicolae, A., J. Wahrheit, Y. Nonnenmacher, C. Weyler and E. Heinzle (2015). "Identification of active elementary flux modes in mitochondria using selectively permeabilized CHO cells." Metab Eng **32**: 95-105.
- Norfolk & Norwich University Hospital, O. a. T. d. (2015). Patient management - PJI.
- Nuxoll, A. S., S. M. Halouska, M. R. Sadykov, M. L. Hanke, K. W. Bayles, T. Kielian, R. Powers and P. D. Fey (2012). "CcpA regulates arginine biosynthesis in *Staphylococcus aureus* through repression of proline catabolism." PLoS Pathog **8**(11): e1003033.
- O'Gara, J. P. (2007). "ica and beyond: biofilm mechanisms and regulation in *Staphylococcus epidermidis* and *Staphylococcus aureus*." FEMS Microbiol Lett **270**(2): 179-188.
- Oberhardt, M. A., B. Ø. Palsson and J. A. Papin (2009). "Applications of genome-scale metabolic reconstructions." Molecular systems biology **5**(1): 320.
- Oduwole, K. O., D. C. Molony, R. J. Walls, S. P. Bashir and K. J. Mulhall (2010). "Increasing financial burden of revision total knee arthroplasty." Knee Surg Sports Traumatol Arthrosc **18**(7): 945-948.
- Onoue, Y. and M. Mori (1997). "Amino acid requirements for the growth and enterotoxin production by *Staphylococcus aureus* in chemically defined media." International Journal of Food Microbiology **36**(1): 77-82.
- Osmon, D. R., E. F. Berbari, A. R. Berendt, D. Lew, W. Zimmerli, J. M. Steckelberg, N. Rao, A. Hanssen and W. R. Wilson (2012). "Diagnosis and management of prosthetic joint infection: clinical practice guidelines by the Infectious Diseases Society of America." Clinical Infectious Diseases: cis803.
- Otto, M. (2008). "Staphylococcal Biofilms." Current topics in microbiology and immunology **322**: 207-228.
- Otto, M. (2009). "*Staphylococcus epidermidis*--the 'accidental' pathogen." Nat Rev Microbiol **7**(8): 555-567.

- Page, A. J., C. A. Cummins, J. A. Keane, J. Parkhill, M. Fookes, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden and D. Falush (2015). "Roary: rapid large-scale prokaryote pan genome analysis." Bioinformatics **31**(22): 3691-3693.
- Parvizi, J., B. Zmistowski, E. F. Berbari, T. W. Bauer, B. D. Springer, C. J. Della Valle, K. L. Garvin, M. A. Mont, M. D. Wongworawat and C. G. Zalavras (2011). "New definition for periprosthetic joint infection: from the Workgroup of the Musculoskeletal Infection Society." Clin Orthop Relat Res **469**(11): 2992-2994.
- Pathology Armony group, U. (2011). Clinical Biochemistry Outcomes. **January**.
- Pedroza-Davila, U., C. Uribe-Alvarez, L. Morales-Garcia, E. Espinoza-Simon, O. Mendez-Romero, A. Muhlia-Almazan, N. Chiquete-Felix and S. Uribe-Carvajal (2020). "Metabolism, ATP production and biofilm generation by *Staphylococcus epidermidis* in either respiratory or fermentative conditions." AMB Express **10**(1): 31.
- Pereira, A. T. (1962). "Coagulase-negative strains of *staphylococcus* possessing antigen 51 as agents of urinary infection." Journal of clinical pathology **15**(3): 252-253.
- Pfeiffer, T., I. Sanchez-Valdenebro, J. C. Nuno, F. Montero and S. Schuster (1999). "METATOOL: for studying metabolic networks." Bioinformatics **15**.
- Phalak, P., J. Chen, R. P. Carlson and M. A. Henson (2016). "Metabolic modeling of a chronic wound biofilm consortium predicts spatial partitioning of bacterial species." BMC Systems Biology **10**(1): 90.
- Poolman, M. (2006). "ScrumPy: metabolic modelling with Python." IEE Proceedings-Systems Biology **153**(5): 375-378.
- Poolman, M. G., S. Kundu, R. Shaw and D. A. Fell (2013). "Responses to light intensity in a genome-scale model of rice metabolism." Plant Physiol **162**(2): 1060-1072.
- Poolman, M. G., L. Miguet, L. J. Sweetlove and D. A. Fell (2009). "A genome-scale metabolic model of *Arabidopsis* and some of its properties." Plant Physiol **151**.
- Poolman, M. G., L. Miguet, L. J. Sweetlove and D. A. Fell (2009). "A genome-scale metabolic model of *Arabidopsis* and some of its properties." Plant Physiol **151**(3): 1570-1581.
- Poolman, M. G., C. Sebu, M. K. Pidcock and D. A. Fell (2007). "Modular decomposition of metabolic systems via null-space analysis." Journal of theoretical biology **249**(4): 691-705.
- Price, N. D., J. L. Reed and B. O. Palsson (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." Nat Rev Microbiol **2**.
- Proctor, R. A., J. M. Balwit and O. Vesga (1994). "Variant subpopulations of *Staphylococcus aureus* as cause of persistent and recurrent infections." Infect Agents Dis **3**(6): 302-312.
- Public Health England, T. S. U., Microbiology Services, PHE (2016). "B 44 Investigation of orthopaedic implant associated infections." UK Standards for Microbiology Investigations(1).
- Raghunathan, A., J. Reed, S. Shin, B. Palsson and S. Daefler (2009). "Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction." BMC Syst Biol **3**: 38.
- Ravikrishnan, A. and K. Raman (2015). "Critical assessment of genome-scale metabolic networks: the need for a unified standard." Brief Bioinform **16**(6): 1057-1068.
- Reitzer, L. (2003). "Nitrogen assimilation and global regulation in *Escherichia coli*." Annu Rev Microbiol **57**: 155-176.
- Resch, A., S. Leicht, M. Saric, L. Pasztor, A. Jakob, F. Gotz and A. Nordheim (2006). "Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling." Proteomics **6**(6): 1867-1877.
- Resch, A., R. Rosenstein, C. Nerz and F. Götz (2005). "Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions." Applied and environmental microbiology **71**(5): 2663-2676.

- Robinson, D. A., A. B. Monk, J. E. Cooper, E. J. Feil and M. C. Enright (2005). "Evolutionary Genetics of the Accessory Gene Regulator (<em>agr</em>) Locus in <em>Staphylococcus aureus</em>." Journal of Bacteriology **187**(24): 8312-8321.
- Rohde, H., S. Frankenberger, U. Zahringer and D. Mack (2010). "Structure, function and contribution of polysaccharide intercellular adhesin (PIA) to Staphylococcus epidermidis biofilm formation and pathogenesis of biomaterial-associated infections." Eur J Cell Biol **89**(1): 103-111.
- Rudin, L., J.-E. Sjöström, M. Lindberg and L. Philipson (1974). "Factors Affecting Competence for Transformation in Staphylococcus aureus." Journal of Bacteriology **118**(1): 155-164.
- Rudin, L., J. E. Sjostrom, M. Lindberg and L. Philipson (1974). "Factors affecting competence for transformation in Staphylococcus aureus." J Bacteriol **118**.
- Sadovskaya, I., E. Vinogradov, J. Li and S. Jabbouri (2004). "Structural elucidation of the extracellular and cell-wall teichoic acids of Staphylococcus epidermidis RP62A, a reference biofilm-positive strain." Carbohydr Res **339**(8): 1467-1473.
- Sadykov, M. R., T. Hartmann, T. A. Mattes, M. Hiatt, N. J. Jann, Y. Zhu, N. Ledala, R. Landmann, M. Herrmann, H. Rohde, M. Bischoff and G. A. Somerville (2011). "CcpA coordinates central metabolism and biofilm formation in Staphylococcus epidermidis." Microbiology **157**(Pt 12): 3458-3468.
- Sadykov, M. R., M. E. Olson, S. Halouska, Y. Zhu, P. D. Fey, R. Powers and G. A. Somerville (2008). "Tricarboxylic acid cycle-dependent regulation of Staphylococcus epidermidis polysaccharide intercellular adhesin synthesis." J Bacteriol **190**(23): 7621-7632.
- Sadykov, M. R., V. C. Thomas, D. D. Marshall, C. J. Wenstrom, D. E. Moormeier, T. J. Widhelm, A. S. Nuxoll, R. Powers and K. W. Bayles (2013). "Inactivation of the Pta-AckA pathway causes cell death in Staphylococcus aureus." J Bacteriol **195**(13): 3035-3044.
- Sadykov, M. R., B. Zhang, S. Halouska, J. L. Nelson, L. W. Kreimer, Y. Zhu, R. Powers and G. A. Somerville (2010). "Using NMR metabolomics to investigate tricarboxylic acid cycle-dependent signal transduction in Staphylococcus epidermidis." J Biol Chem **285**(47): 36616-36624.
- Sasarman, A., P. Purvis and V. Portelance (1974). "Role of menaquinone in nitrate respiration in Staphylococcus aureus." J Bacteriol **117**(2): 911-913.
- Schepens, D., R. Carlson, J. Heys, A. Beck and T. Gedeon (2019). "Role of resource allocation and transport in emergence of cross-feeding in microbial consortia." Journal of Theoretical Biology **467**.
- Schilling, C. H. and B. O. Palsson (2000). "Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis." J Theor Biol **203**.
- Schomburg, I., O. Hofmann, C. Baensch, A. Chang and D. Schomburg (2000). "Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine." Gene Function & Disease **1**(3-4): 109-118.
- Schuster, S., T. Dandekar and D. A. Fell (1999). "Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering." Trends in biotechnology **17**(2): 53-60.
- Schuster, S. and D. Fell (2007). "Modeling and simulating metabolic networks." Bioinformatics-from genomes to therapies: 755-805.
- Schuster, S., T. Pfeiffer and D. A. Fell (2008). "Is maximization of molar yield in metabolic networks favoured by evolution?" J Theor Biol **252**(3): 497-504.
- Sebastian, S., R. Malhotra and B. Dhawan (2018). "Prosthetic joint infection: A major threat to successful total joint arthroplasty." Indian J Med Microbiol **36**(4): 475-487.

- Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." Bioinformatics **30**(14): 2068-2069.
- Shapiro, B. A. (1995). "Temperature correction of blood gas values." Respir Care Clin N Am **1**(1): 69-76.
- Shlomi, T., Y. Eisenberg, R. Sharan and E. Ruppin (2007). "A genome-scale computational study of the interplay between transcriptional regulation and metabolism." Molecular systems biology **3**(1): 101.
- Singh, D., R. Carlson, D. Fell and M. Poolman (2015). "Modelling metabolism of the diatom *Phaeodactylum tricornutum*." Biochem Soc Trans **43**(6): 1182-1186.
- Sivakanesan, R. and E. A. Dawes (1980). "Anaerobic glucose and serine metabolism in *Staphylococcus epidermidis*." J Gen Microbiol **118**(1): 143-157.
- Somerville, G. A. (2016). Staphylococcus: genetics and physiology. UK, Caister Academic Press.
- Somerville, G. A., C. I. Morgan, J. R. Fitzgerald, D. W. Dorward, L. J. Reitzer and J. M. Musser (2002). "Staphylococcus aureus aconitase inactivation unexpectedly inhibits post-exponential-phase growth and enhances stationary-phase survival." Infection and immunity **70**(11): 6373-6382.
- Somerville, G. A., B. Said-Salim, J. M. Wickman, S. J. Raffel, B. N. Kreiswirth and J. M. Musser (2003). "Correlation of acetate catabolism and growth yield in *Staphylococcus aureus*: implications for host-pathogen interactions." Infect Immun **71**(8): 4724-4732.
- Somerville, G. A., B. Said-Salim, J. M. Wickman, S. J. Raffel, B. N. Kreiswirth and J. M. Musser (2003). "Correlation of acetate catabolism and growth yield in *Staphylococcus aureus*: implications for host-pathogen interactions." Infect Immun **71**.
- Stein, W. H. and S. Moore (1954). "The free amino acids of human blood plasma." J Biol Chem **211**(2): 915-926.
- Stepanovic, S., D. Vukovic, V. Hola, G. Di Bonaventura, S. Djukic, I. Cirkovic and F. Ruzicka (2007). "Quantification of biofilm in microtiter plates: overview of testing conditions and practical recommendations for assessment of biofilm production by staphylococci." Apmis **115**(8): 891-899.
- Stephanopoulos, G., Aristidou, A., Nielsen, J. (1998). Metabolic engineering. Principles and methodologies. San Diego, Academic Press.
- Sutherland, I. W. (2001). "The biofilm matrix--an immobilized but dynamic microbial environment." Trends Microbiol **9**(5): 222-227.
- Takahashi, T., I. Satoh and N. Kikuchi (1999). "NOTE." International Journal of Systematic and Evolutionary Microbiology **49**(2): 725-728.
- Tan, L., S. R. Li, B. Jiang, X. M. Hu and S. Li (2018). "Therapeutic Targeting of the *Staphylococcus aureus* Accessory Gene Regulator (*agr*) System." Frontiers in Microbiology **9**(55).
- Tande, A. J. and R. Patel (2014). "Prosthetic joint infection." Clinical microbiology reviews **27**(2): 302-345.
- Tempest, D. W., J. L. Meers and C. M. Brown (1970). "Synthesis of glutamate in *Aerobacter aerogenes* by a hitherto unknown route." Biochem J **117**(2): 405-407.
- Townsend, D. E., A. Kaenjak, R. K. Jayaswal and B. J. Wilkinson (1996). "Proline is biosynthesized from arginine in *Staphylococcus aureus*." Microbiology **142** ( Pt 6): 1491-1497.
- Trinh, C. T., P. Unrean and F. Srienc (2008). "Minimal *Escherichia coli* Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses." Applied and Environmental Microbiology **74**(12): 3634-3643.

- Tsikas, D., I. Fuchs, F. M. Gutzki and J. C. Frolich (1998). "Measurement of nitrite and nitrate in plasma, serum and urine of humans by high-performance liquid chromatography, the Griess assay, chemiluminescence and gas chromatography-mass spectrometry: interferences by biogenic amines and N(G)-nitro-L-arginine analogs." J Chromatogr B Biomed Sci Appl **715**(2): 441-444; discussion 445-448.
- Tynecka, Z., Z. Szczesniak, A. Malm and R. Los (1999). "Energy conservation in aerobically grown *Staphylococcus aureus*." Res Microbiol **150**(8): 555-566.
- Unden, G. and J. Bongaerts (1997). "Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors." Biochim Biophys Acta **1320**(3): 217-234.
- Uribe-Alvarez, C., N. Chiquete-Felix, M. Contreras-Zentella, S. Guerrero-Castillo, A. Pena and S. Uribe-Carvajal (2016). "Staphylococcus epidermidis: metabolic adaptation and biofilm formation in response to different oxygen concentrations." Pathog Dis **74**(1): ftv111.
- Van Dien, S. J. and M. E. Lidstrom (2002). "Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C(3) and C(4) metabolism." Biotechnol Bioeng **78**.
- Varma, A. and B. O. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110." Appl Environ Microbiol **60**.
- Vazquez, A. (2018). Overflow Metabolism: From Yeast to Marathon Runners., Academic Press.
- Villanova, V., A. E. Fortunato, D. Singh, D. D. Bo, M. Conte, T. Obata, J. Jouhet, A. R. Fernie, E. Marechal, A. Falciatore, J. Pagliardini, A. Le Monnier, M. Poolman, G. Curien, D. Petroutsos and G. Finazzi (2017). "Investigating mixotrophic metabolism in the model diatom *Phaeodactylum tricornutum*." Philos Trans R Soc Lond B Biol Sci **372**(1728).
- Von Eiff, C., C. R. Arciola, L. Montanaro, K. Becker and D. Campoccia (2006). "Emerging *Staphylococcus* species as new pathogens in implant infections." The International journal of artificial organs **29**(4): 360-367.
- Vuong, C., J. B. Kidder, E. R. Jacobson, M. Otto, R. A. Proctor and G. A. Somerville (2005). "Staphylococcus epidermidis polysaccharide intercellular adhesin production significantly increases during tricarboxylic acid cycle stress." J Bacteriol **187**(9): 2967-2973.
- Wang, X., J. F. Preston and T. Romeo (2004). "The pgaABCD Locus of *Escherichia coli* Promotes the Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation." Journal of Bacteriology **186**(9): 2724-2734.
- Washburn, R. S., A. Marra, A. P. Bryant, M. Rosenberg and D. R. Gentry (2001). "rho Is Not Essential for Viability or Virulence in *Staphylococcus aureus*." Antimicrob Agents Chemother **45**(4): 1099-1103.
- Weigent, D. A. and E. W. Nester (1976). "Purification and properties of two aromatic aminotransferases in *Bacillus subtilis*." J Biol Chem **251**(22): 6974-6980.
- Wetzel, K. J., D. Bjorge and W. R. Schwan (2011). "Mutational and Transcriptional Analyses of the *Staphylococcus aureus* Low Affinity Proline Transporter OpuD During in vitro Growth and Infection of Murine Tissues." FEMS Immunol Med Microbiol **61**(3): 346-355.
- Wilkinson, B. (1997). *The staphylococci in human disease.* Biology, Churchill Livingstone, NY, USA: 1-38.
- Wimpenny, J., W. Manz and U. Szewzyk (2000). "Heterogeneity in biofilms." FEMS Microbiol Rev **24**(5): 661-671.
- Wimpenny, J. W. and J. P. Coombs (1983). "Penetration of oxygen into bacterial colonies." J Gen Microbiol **129**(4): 1239-1242.

- Wu, K., J. Conly, M. Surette, C. Sibley, S. Elsayed and K. Zhang (2012). "Assessment of virulence diversity of methicillin-resistant *Staphylococcus aureus* strains with a *Drosophila melanogaster* infection model." BMC Microbiol **12**: 274.
- Xiao, Z. and P. Xu (2007). "Acetoin metabolism in bacteria." Critical Reviews In Microbiology **33**(2): 127-140.
- Xue, T., J. Ni, F. Shang, X. Chen and M. Zhang (2015). "Autoinducer-2 increases biofilm formation via an *ica*- and *bhp*-dependent manner in *Staphylococcus epidermidis* RP62A." Microbes and Infection **17**(5): 345-352.
- Yao, Y., D. E. Sturdevant and M. Otto (2005). "Genomewide analysis of gene expression in *Staphylococcus epidermidis* biofilms: insights into the pathophysiology of *S. epidermidis* biofilms and the role of phenol-soluble modulins in formation of biofilms." J Infect Dis **191**(2): 289-298.
- Zapotoczna, M., E. O'Neill and J. P. O'Gara (2016). "Untangling the Diverse and Redundant Mechanisms of *Staphylococcus aureus* Biofilm Formation." PLoS Pathog **12**(7): e1005671.
- Zapotoczna, M., E. O'Neill and J. P. O'Gara (2016). "Untangling the diverse and redundant mechanisms of *Staphylococcus aureus* biofilm formation." PLoS Pathog **12**(7): e1005671.
- Zhang, T., A. Parker, R. Carlson, P. Stewart and I. Klapper (2018). Flux-Balance Based Modeling of Biofilm Communities.
- Zhu, J., E. Dizin, X. Hu, A. S. Wavreille, J. Park and D. Pei (2003). "S-Ribosylhomocysteinase (LuxS) is a mononuclear iron protein." Biochemistry **42**(16): 4717-4726.
- Zhu, W., K. Sieradzki, V. Albrecht, S. McAllister, W. Lin, O. Stuchlik, B. Limbago, J. Pohl and J. K. Rasheed (2015). "Evaluation of the Biotyper MALDI-TOF MS system for identification of *Staphylococcus* species." Journal of microbiological methods **117**: 14-17.
- Zhu, Y., E. C. Weiss, M. Otto, P. D. Fey, M. S. Smeltzer and G. A. Somerville (2007). "Staphylococcus aureus biofilm metabolism and the influence of arginine on polysaccharide intercellular adhesin synthesis, biofilm formation, and pathogenesis." Infect Immun **75**(9): 4219-4226.
- Zwietering, M. H., I. Jongenburger, F. M. Rombouts and K. van 't Riet (1990). "Modeling of the bacterial growth curve." Appl Environ Microbiol **56**(6): 1875-1881.

## 9 Appendices

### 9.1 Appendix A: additional material

Content of the additional directory Code\_TDC\_2020

#### 9.1.1 Python modules for model construction

##### 9.1.1.1 Model construction

Model construction was done by calling the top level module BuildSepi.py. This module determines the PGDB database from which the organism specific reactions are to be imported and calls the modules UsrBuildOrg.py, BuildOrg.py and CompartmentDic.py. These contain the code needed to build the GSM (UsrBuildOrg.py), apply the changes specified in the stoichiometry and directionality correction files, remove unwanted reactions and metabolites while renaming other metabolites to avoid name inconsistencies (BuildOrg.py) and assign specific reactions to their corresponding compartments (CompartmentDic.py). In this case, the model considers the cell cytosol to be the sole system compartment. Sepi.spy is the main module to be called when loading the model object. It imports other modules spread in multiple files as described in Chapter 2, Section 2.4.2. The subdirectory Tools contains a module to generate a general LP object (BuildLP.py) and to generate a dictionary of fluxes for biomass exporters corresponding to the proportions of biomass components experimentally determined in the biomass (Chapter 2, Section 2.4.3.4) to be used as constrains for biomass production (SepiBiomass.py).

Modules to be called when loading the model object for the specific analysis described in each chapter:

- Sepi\_MinMed.spy: Chapter 2 – model curation; Chapter 3 – model analysis for ATP production from Glc; production of planktonic biomass under a range of environmental conditions.
- Sepi\_MinMedGLT.spy: Chapter 3 – model analysis for ATP production from Glt.
- Sepi\_MM.spy: Chapter 4 – model analysis for defining the effect of single amino acid deprivation on biomass production. Chapter 5 – model analysis for the study of N and amino acids utilisation for biomass production.
- Sepi\_synovial.spy: Chapter 6 – model analysis for ATP or biomass production in synovial fluid while responding to changes in the O<sub>2</sub> concentration.

##### 9.1.1.2 Deleted reactions and metabolites

List of deleted reactions and metabolites. Python file (Unwanted.py).

### ***9.1.1.3 Corrected reactions and metabolites***

List of reactions stoichiometry and reversibility corrections applied and metabolites renamed. Python files (Corr\_Stoich.spy, Corr\_Revers.spy and Substitutes.py).

## **9.1.2 Python modules for constructive analysis of the model**

### ***9.1.2.1 General properties of the model***

Methods used in Chapters 2 and 3 to define general model properties such as number of reactions, metabolites, gene associations, reaction subsets, etc. Python file (ModelGeneralProperties.py).

Functions:

- ModelProperty: determines general properties of the model
- ElementaryModesETC: identification of elementary modes through the ETC

### ***9.1.2.2 General validation of the model***

Methods used to perform general validation of the model such as testing the model for energy, redox and mass conservation, reactions atomic balance or stoichiometry consistency in Chapter 2. Python file (ModelGeneralValidation.py).

Functions:

- AtomiCheck: atomic balance of reactions
- StoiCons: material consistency of the model
- MassCons: model-wide mass conservation
- ATPCons: model-wide energy conservation
- NADHCons; NADPHCons: model-wide redox conservation

### ***9.1.2.3 ETC stand-alone module***

Module including the ETC.spy module (reactions involved in generation of the proton motive force) plus any other reactions involved in the electron transport chain so it can be analysed in isolation in Chapter 3. Python file (ETC\_TopLevel.spy).

### ***9.1.2.4 Feasibility of production of single biomass components***

Methods used to check feasibility of production of individual biomass components in Chapters 2 and 4. Python file (BiomassProd.py).

Functions:

- BiomassProd.py (fx CheckProds): feasibility of production of biomass components one by one.
- Check: production of one unit of a compound of interest.
- CheckProds: feasibility of production of each single biomass component independently.

- CheckCandReacsREV: identification of candidate reactions for thermodynamic re-definition.

### ***9.1.2.5 Feasibility of biomass production***

Methods used to check feasibility of biomass production while meeting ATP cell maintenance requirements in Chapter 4. Python file (ProdUnitsBM.py).

Functions:

- ProdUnitBM: feasibility of production of one unit of cell biomass while meeting the ATP cell maintenance requirements.

### ***9.1.2.6 ATP production***

Methods used to check feasibility and efficiency of generation of energy as ATP from different C sources in Chapter 3. Python files (ATPprod.py and ATPprodFromBPs.py).

Functions:

- ATPprodResponse: ATP production from various C sources.
- ATPprodFromAc: ATP production from Ac.
- PrintValsProdATP: ATP production minimising export of by-products leading to media acidification.
- ATPprodFromBPs: ATP production when metabolic by-products are available as C sources.

### ***9.1.2.7 Essentiality of media components***

Methods used to check essentiality of media components for biomass production in in Chapter 2. Python file (Essential.py).

Functions:

- EssentialComps: essentiality of media components.

## **9.1.3 Python modules for functional analysis of the model**

### ***9.1.3.1 Essentiality of amino acids and effect of amino acid deprivation on biomass production.***

Methods used to check essentiality of amino acids for biomass production and the effect of single amino acid removal in the objective value (total net flux through the system) and the demand of Glc in Chapters 4 and 5. Python file (AAessential.py).

Functions:

- PrintValsStdBMP: production of biomass in a specific medium (MM minimal medium etc).
- PrintValsBMPinNH<sub>4</sub><sup>+</sup>: production of biomass with NH<sub>4</sub><sup>+</sup> as sole N source.

- EffectAArmBMP: effect of single amino acids removal on the objective value and Glc demand of solutions for biomass production with respect to the solution obtained in the unmodified media.
- PrintValsUponAArm: objective value and Glc demand of solutions for biomass production upon single amino acid removal.

### ***9.1.3.2 Utilisation for biomass production and assimilation of inorganic N.***

Methods used to investigate differential utilisation of amino acids for biomass production and the assimilation of inorganic N in Chapters 4 and 5. Python files (AAutilN.py and AAutilC.py).

Functions:

- UtilAA: overall uptake/excretion values for amino acids during synthesis of biomass.
- ContribNuptakeAAs: contribution of each amino acid to the total N uptake for biomass production.
- ContribCuptakeAAs: contribution of each amino acid to the total C uptake for biomass production.
- EffectBMPinSingleAA: effect of utilising single amino acids for biomass production was calculated as the percentage of increase in the objective value with respect of the objective value
- AAccheckProds: amino acids that can be produced when a single amino acid is utilised at a time as N source.
- UtilSingleAA\_C: potential of each single amino acid to be used as sole C source.
- Nassimilation: synthesis of Glt from Glc and with  $\text{NH}_4^+$  as sole N source.

### ***9.1.3.3 Oxygen scan***

Methods used to perform  $\text{O}_2$  scan analysis for ATP production and for biomass production in Chapter 6. Python files (O2ScanATPonly.py, O2Scan.py).

Functions:

- O2ScanATPonly.py: responses to ATP production upon variation in the  $\text{O}_2$  concentration.
- O2Scan.py: responses to biomass production upon variation in the  $\text{O}_2$  concentration.

## ***9.2 Appendix B: biomass composition***

Biomass composition for *S. epidermidis* used to set up the constraints for LP-based analysis of the model involving biomass production in this project.

**Table 9-1 Biomass composition for *S. epidermidis* derived from data corresponding to *S. aureus* and modified on basis to biochemical data currently available for *S. epidermidis*.**

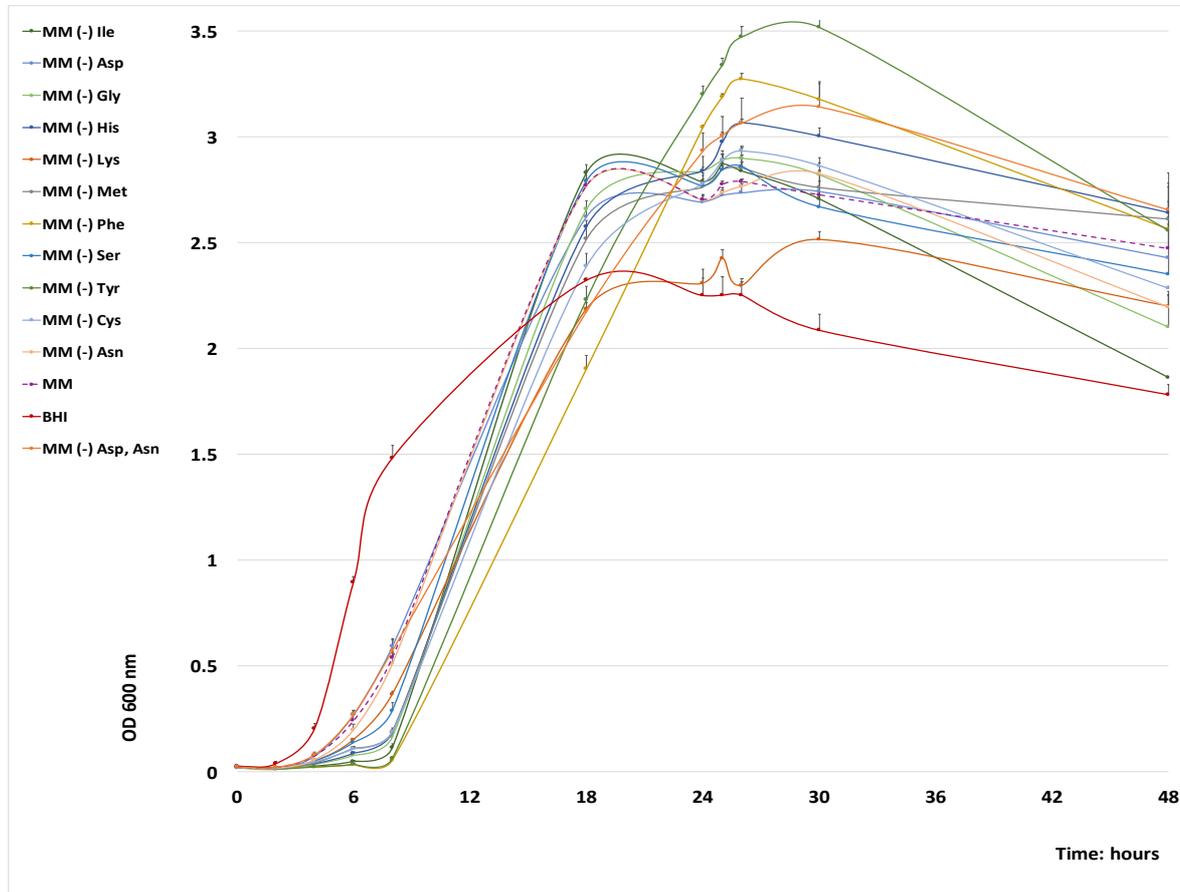
<b>Component</b>	<b>Percentage per unit of biomass</b>	<b>Concentration (mmol/gDW/h)</b>
<b>Protein</b>	<b>43</b>	-
Ala	4.40	0.212
Arg	4.49	0.111
Asn	5.31	0.173
Asp	8.02	0.261
Cys	0.54	0.0191
Gln	6.80	0.200
Glt	6.78	0.199
Gly	3.31	0.190
His	2.65	0.0734
Ile	8.21	0.269
Leu	8.60	0.282
Lys	8.06	0.235
Met	2.90	0.0837
Phe	5.28	0.137
Pro	3.11	0.116
Ser	4.83	0.198
Thr	4.95	0.179
Tyr	5.01	0.119
Val	5.65	0.207
<b>DNA</b>	<b>3</b>	-
DATP	32.9	0.0203
DCTP	15.3	0.0990
DGTP	16.6	0.0990
DTTP	32.3	0.0203
<b>RNA</b>	<b>12</b>	-

ATP	25.5	0.0607
CTP	23.8	0.0595
GTP	25.7	0.0595
UTP	24.3	0.0607
<b>Cell membrane</b>	<b>7</b>	-
Diacylglycerol	16.0	0.0186
1-phosphatidylglycerol	48.1	0.0446
Cardiolipin	11.3	0.00559
Lipoteichoic acid type I	6.83	0.00335
Glc2-DAG (diglucosyl- diacylglycerol)	8.86	0.00669
MGlcDG (3-D-glucosyl-1,2- diacylglycerol)	1.22	0.00112
Menaquinone	7.63	0.00744
<b>Cell wall</b>	<b>24</b>	-
Peptidoglycan with D,D cross- link (CPD-12230)	23.4	0.0101
Peptidoglycan-wall teichoic acid complex	76.6	0.0623
<b>Pool of solutes</b>	<b>1.1</b>	-
Glycogen	86.5	0.0528
AcCoA	0.267	3.65e-05
SucCoA	0.0173	2.21e-06
CoA	0.307	4.42e-05
FAD	0.527	7.41e-05
NAD	9.54	0.00158
NADH	0.220	3.65e-05
NADP	0.647	9.62e-05
NADPH	1.99	2.95e-05

This biomass composition was defined as described in Chapter 2, Section 2.4.3.4. A growth rate of 1g/h was assumed (Chapter 3, Section 3.3.4).

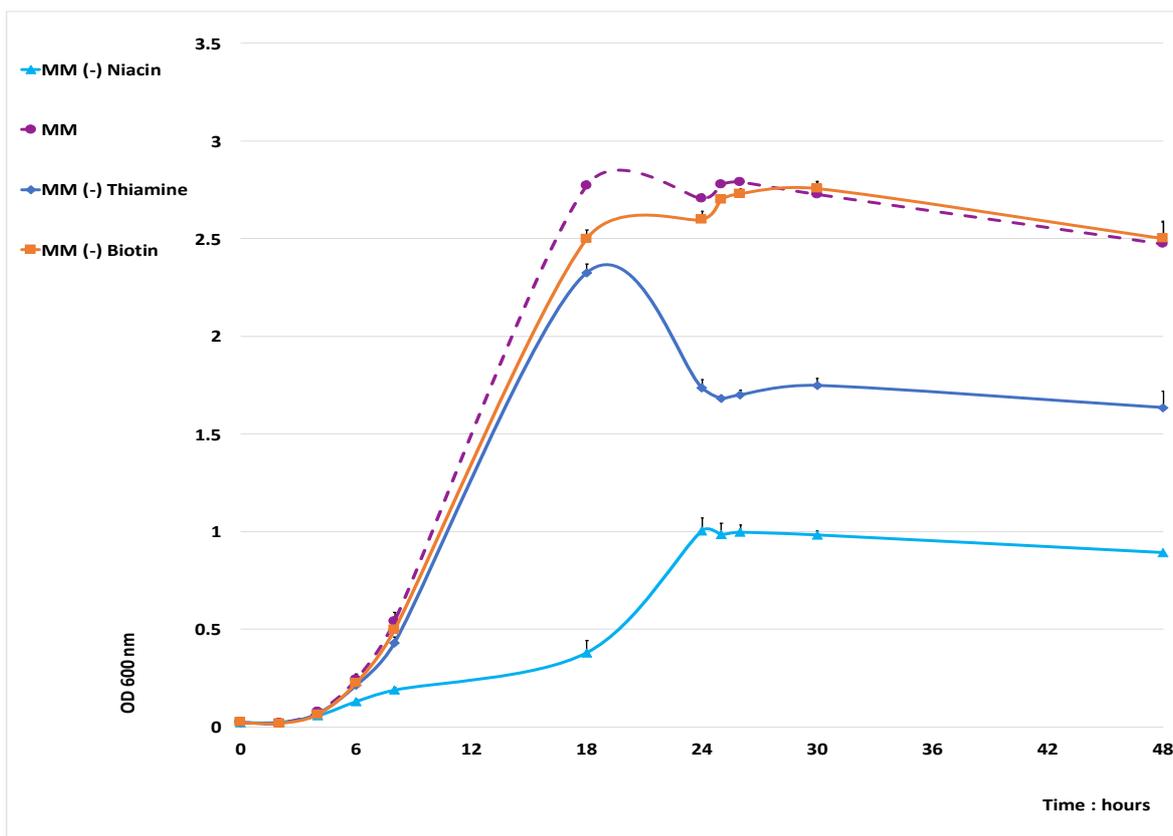
## 9.3 Appendix C: experimental data

### 9.3.1 Minimal growth requirements



**Figure 9-1 Growth curves for *S. epidermidis* RP62A in all modified MM medium samples presenting no apparent delay in growth in comparison to cultures in standard MM medium (purple)**

Note that the growth curve for cultures in BHI medium (red) was included for reference of growth in a rich medium. Legend: BHI = BHI medium; MM = standard MM medium; MM (-) 'compound/s name' = MM medium without 'compound/s name'. Error bars = SEM



**Figure 9-2 Growth curves for *S. epidermidis* RP62A in modified MM<sup>-</sup> medium samples without single vitamins compared to cultures in standard MM medium (purple)**

Legend: purple dashed line (MM) = standard MM medium; light blue line (MM (-) niacin) = MM medium without niacin; dark blue line (MM (-) thiamine) = MM medium without thiamine; orange line (MM (-) biotin) = MM medium without biotin. Each data point corresponds to the mean  $A_{600}$  value of three independent biological replicates. Error bars = SEM.

**Table 9-2 Comparison between the biosynthetic potential for vitamins encoded in the genome of *S. epidermidis* RP62A, vitamin auxotrophies reported experimentally and their essentiality according to the GSM after curation**

Vitamin	Absence of biosynthetic genes in the RP62A	Auxotrophy reported by Hussain <i>et al.</i> (1991)	Auxotrophy experimentally determined in this study	Essentiality according to model after curation	Proposed explanation for discrepancies between experimental datasets
Biotin	No	Included in medium by default.	No	NA (not included in biomass).	Normal growth observed without biotin and the presence of a biosynthetic pathway indicates lack of essentiality.
Niacin	Yes	Included in medium by default.	No, but its absence delays growth.	Yes	Since no biosynthetic pathway is present, growth must be due to carry over of small amounts of niacin in the inoculum or to niacin uptake after cell-lysis.
Thiamine	Yes	Included in medium by default.	No	NA (not included in biomass).	Since no biosynthetic pathway is present, growth must be due to carry over of small amounts of thiamine in the inoculum or to its uptake after cell-lysis.

RP62A seems to lack the genetic potential for *de novo* synthesis of the vitamins niacin and thiamine (Table 9-2), and, in consequence, the model identifies niacin as essential for growth, while thiamine is not directly involved in biomass production by the system. This dependency on niacin matches published data on the subject (Hussain *et al.* 1991; Heinemann *et al.* 2005; Lee *et al.* 2009; Bosi *et al.* 2016). However, although severely delayed and reduced, experimental results showed growth in niacin-deprived cultures (Figure 9-2).

The reason for this discrepancy is unclear: it is possible that, since cofactors are required in very small amounts, a potential residual carry over of these vitamins during inoculum preparation (Chapter 4, Section 4.2.1.2) could have been enough to allow some growth. Alternatively, they could potentially have been uptaken from the media after cell-lysis.

### 9.3.2 Impact of amino acid deprivation on biofilm formation

Table 9-3 Relative levels of biofilm formation detected for RP62A cultures growing in different media cultures growing in different media and variations observed with respect to cultures in the standard MM medium.

Media	Relative level of biofilm formation (%)	Variation in biofilm formation with respect to the standard MM medium (%)	Cell growth at time 48 hours ( $A_{600}$ )
Standard MM medium	92.5 ± 5.82	-	2.47 ± 0.28
MM medium (-) Glt	4.93 ± 2.99	-94.7 ± 5.39	2.65 ± 0.37
MM medium (-) Val	9.86 ± 2.08	-89.3 ± 1.92	1.17 ± 0.62
MM medium (-)Pro	14.0 ± 0.74	-84.9 ± 0.68	0.11 ± 0.00
MM medium (-) Leu	14.5 ± 8.61	-84.3 ± 7.96	2.91 ± 0.12
MM medium (-) Arg	24.1 ± 9.04	-73.9 ± 8.36	1.97 ± 0.33
MM medium (-) His	28.5 ± 25.2	-69.2 ± 23.3	2.64 ± 0.17
MM medium (-) Phe	58.5 ± 10.6	-36.8 ± 9.81	2.56 ± 0.03
MM medium (-) Gly	71.7 ± 11.0	-22.5 ± 10.2	2.10 ± 0.31
MM medium (-) Trp	84.0 ± 2.77	-9.20 ± 2.56	2.79 ± 0.13
MM medium (-) Cys	85.7 ± 5.41	-7.31 ± 5.00	2.28 ± 0.02
MM medium (-) Lys	88.4 ± 11.5	-4.43 ± 10.64	2.20 ± 0.20
MM medium (-) Tyr	89.8 ± 4.93	-2.92 ± 4.56	2.56 ± 0.23
MM medium (-) Ser	90.3 ± 3.11	-2.43 ± 2.88	2.35 ± 0.39
MM medium (-) Asn	94.4 ± 9.37	+2.06 ± 8.66	2.20 ± 0.13
MM medium (-) Met	94.8 ± 4.00	+2.43 ± 3.70	2.61 ± 0.45
MM medium (-) Asp	95.0 ± 14.0	+2.65 ± 13.0	2.43 ± 0.13
MM medium (-) Ile	95.9 ± 10.9	+3.65 ± 10.0	1.86 ± 0.11
MM medium (-) Ala	96.2 ± 3.00	+3.99 ± 2.78	1.60 ± 0.21
MM medium (-) Thr	100 ± 3.75	+8.01 ± 3.47	2.58 ± 0.58

Values of  $A_{600}$  obtained before staining of the biofilm biomass are included as a reference for cell growth in these samples. Error values = SD.

## 9.4 Appendix D: model refinement

### 9.4.1 Comparison between experimental data and *in silico* results on the essentiality of media components

As a result of the work described in Chapter 4, Section 4.2.3, several reactions were introduced, modified or removed from the system:

#### 9.4.1.1 Biosynthesis of amino acids

##### 1. Biosynthesis of Arg:

**Experimental data:** current experimental results suggest that this amino acid is not essential although its absence delayed growth notably. Previous work considered RP62A to be auxotrophic for Arg (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** Arg *de novo* synthesis from glutamine and/or Glt is completed in the genome of RP62A. The model can synthesize Arg. These reactions lead to synthesis of Arg via ornithine (through the urea cycle). However, the urea cycle is broken. The enzyme arginase is missing, and so, the direct hydrolysis of Arg into urea and ornithine cannot occur. This could potentially affect regeneration of ornithine that is in turn used to obtain other compounds such as Pro. However, ornithine can still be obtained from Arg via citrulline (EC 3.5.3.6 and EC 2.1.3.3), in a process that also produces Carbamoyl-P.

**Model refinement:** the reaction catalysed by the enzyme arginase (EC 3.5.3.1) had been originally included into the PGDBs for RP62A as a result of the automatic gap-filling process carried out by the Pathway Tools software (and therefore was present in the model). For accuracy, this reaction was removed, and the urea cycle broken in consequence. Even though this does not prevent Arg synthesis *in-silico* it might have unpredicted biological effects.

##### 2. Biosynthesis of Trp:

**Experimental results:** our experimental results suggest that this amino acid is not essential although its absence delayed growth severely. Previous work considered RP62A to be auxotrophic for Trp (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** a synthetic pathway for *de novo* synthesis of Trp from chorismate is present in RP62A. The model is in consequence able to synthesise this amino acid.

**Model refinement:** none. Since a biosynthetic path is present and no significant mutations

of the enzymes involved were found, lack of growth without Trp (Hussain *et al.* 1991) is assumed to be due to regulatory processes.

### 3. Biosynthesis of Val:

**Experimental results:** our experimental results suggest that this amino acid is not essential although its absence delayed growth drastically. Previous work considered RP62A to be auxotrophic for Val (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** a synthetic pathway for *de novo* synthesis of Val from pyruvate is present in RP62A. The model can consequently synthesise Val.

**Model refinement:** none. Since a biosynthetic path is present and no significant mutations of the enzymes involved were found, lack of growth without Val (Hussain *et al.* 1991) is assumed to be due to regulatory processes.

### 4. Biosynthesis of Pro:

**Experimental results:** our experimental results suggest that RP62A is auxotrophic for Pro. However, Husain *et al.* reported Pro deprivation to have no effect on growth (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** *S. epidermidis* RP62A is lacking two enzymes involved in the biosynthesis of Pro from Glt via glutamyl-P (EC 2.7.2.11 and EC 1.2.1.41). However, the presence of three other enzymes (EC 2.6.1.13, EC 1.5.1.2 and EC 4.3.1.12) shown by analysis of the genome enables a bypass that could be used to obtain Pro from ornithine (Lee *et al.* 2009) which can be obtained from Arg via citrulline or even from Glt via a much longer process. As a result of this, the GSM theoretically predicts Pro production and biomass production without and external source of Pro. Similar observations for *S. aureus* have previously been described in the literature when analysing minimal media for staphylococci (Lee *et al.* 2009). However, in this case, the authors concluded that Pro should be added to the minimal medium in other to support growth.

**Model refinement:** none. The fact that previous experimental data did not show an auxotrophy for Pro together with the presence of a biosynthetic bypass through ornithine and the fact that no significant mutations were found on the laboratory strain suggest that lack of growth without Pro must be due to either cells not being able to obtain enough ornithine for its synthesis from the MM medium without Pro or to regulatory processes.

### 5. Biosynthesis of Cys:

**Experimental results:** our experimental results suggest that the absence of this amino acid has no drastic effect on growth. Previous work considered RP62A to be auxotrophic for Cys (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** Cys biosynthetic pathways from Ser and Met (via homoCys) are present in RP62A. However, the model initially predicted an auxotrophy for Cys.

**Model refinement:** after investigation, an exporter for the compound autoinducer 2 (UI CPD-10774) was included. This compound is involved in quorum sensing and produced in the S-adenosyl-L-Met cycle (MetaCyc 22.6 (Caspi *et al.* 2014)), together with homoCys. This allowed Cys biosynthesis from Met but still did not allow biomass production without exogenous Cys. The model was not able to obtain Cys from Ser utilising EC 2.3.1.30 and EC 2.5.1.47: sulphide is needed in order to obtain Cys through this route, however, two reactions respectively catalysed by EC 1.8.4.8 and EC 1.8.1.2 are involved in sulphide production from sulphate, and were dead, which prevented production of this compound. This was due to 3'-5'-ADP being produced during the assimilatory sulphate reduction pathway by EC 1.8.4.8 and accumulating in the system, thus breaking the steady state assumption. This was fixed by including a reaction that allowed hydrolysis of 3'-5'-ADP to AMP and phosphate (EC 3.1.3.7). The corresponding gene associated with EC 3.1.3.7 in RP62A (SerP1267) was found in BioCyc and KEGG. ACT was used to identify its presence in the genome of the lab strain. No obvious mutations were found on these sequences and a Pfam domain for the DHH phosphatase family (phosphoesterases) was identified, which justified inclusion of the reaction in the model. This allowed Cys biosynthesis from Ser, and finally, biomass production without external Cys.

## 6. Biosynthesis of Met:

**Experimental results:** our experimental results are in agreement with previous findings and suggest that this is not an essential amino acid for growth (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** biosynthetic pathways for Met from Ser and Cys (via homoCys) are present in RP62A. However, the model initially predicted an auxotrophy for Met.

**Model refinement:** the absence of two reactions catalysed by cystathione gamma-synthase (EC 2.5.1.48) was breaking the Met biosynthetic pathway from homoCys. After investigation, identification of genes associated with this enzyme on the strain's genome (SerP0037) justified the inclusion of both reactions in the model. This allowed synthesis of Met. A set of three extra reactions catalysed by EC 2.5.1.48 and EC 4.4.1.8 (SerP0036) was included for accuracy. These reactions were present in the KEGG pathway database for RP62A but missing on BioCyc.

## 7. Biosynthesis of Asp:

**Experimental results:** our experimental results are in agreement with previous findings and suggest that this is not an essential amino acid for growth (Hussain *et al.* 1991). No effect on cell growth was observed when Asp and Asn were removed simultaneously.

**Biosynthetic potential and model behaviour:** according to BioCyc and KEGG, *de novo* synthesis of Asp from Glt and oxaloacetate is broken in RP62A due to the absence of the enzyme Asp aminotransferase (EC 2.6.1.1). However, conversion of Asn into Asp is possible thanks to the enzyme asparaginase (EC 3.5.1.1), with the corresponding reaction being present in the model. Initially, the model could produce Asp as long as exogenous Asn was provided. Since experimental results suggested that Asp biosynthesis without Asn must also be possible, further investigation was performed: the BioCyc database for RP62A showed the presence of a gene (SerP2159) which had been automatically assigned through protein homology to EC 2.6.1.1. However, this reaction was missing in the model. Study of the amino acid sequence of SerP2159 in the lab strain predicted its association with a Pfam domain: aminotransferase class-III. This, together with the experimental evidence justified inclusion of the corresponding reaction catalysed by Asp aminotransferase, ultimately allowing *de novo* synthesis of Asp.

## 8. Biosynthesis of Asn:

**Experimental results:** our experimental results are in agreement with previous findings and suggest that this is not an essential amino acid for growth (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** the model initially predicted an auxotrophy for this amino acid. This is due to RP62A lacking the genes encoding common Asn synthases: EC 6.3.1.1 (in *E. coli* and *S. aureus*) or EC 6.3.5.4 (in *E. coli*). An alternative route for Asn synthesis is the conversion of Asp to Asn via a tRNA-dependent transamidation mechanism (Curnow *et al.* 1998), for which the corresponding hypothetical enzymes have been identified in RP62A (EC 6.1.1.22, 6.3.5.6 and 3.1.1.29). Their associated reactions were included in BioCyc but had been initially removed from the model in order to reduce the number of compounds with undefined empirical formulae and therefore simplify detection of possible stoichiometric inconsistencies in the system.

**Model refinement:** experimental data proved that this strain can obtain Asn from Asp. The reactions catalysed by EC 6.1.1.22, 6.3.5.6 and 3.1.1.29 were included in the model. The model can now perform tRNA-dependent synthesis of Asn from Asp.

## 9. Biosynthesis of Phe:

Note that the biosynthetic pathway for Phe had previously been curated and completed prior to the beginning of these experiments.

**Experimental results:** our experimental results are in agreement with previous findings and suggest that this is not an essential amino acid for growth (Hussain *et al.* 1991).

**Biosynthetic potential and model behaviour:** the biosynthetic pathway for Phe and Tyr

production from chorismate is completed in RP62A. The enzymes EC 2.6.1.9 and EC 2.6.1.57, both present in the model and encoded by SerP0387 (hisC), catalise the main transamination reactions involved in the synthesis of Tyr and Phe from phenylpyruvate.

**Model refinement:** with the corresponding reactions already present in the model and involving Tyr and histidinol phosphate as substrates/products (the reactions are reversible), another reversible reaction was included utilizing Phe, which had been previously reported as another valid substrate of these transaminases by Weigen and Nester in 1976 (Weigent *et al.* 1976). This allowed *in-silico* biosynthesis of Phe from phenylpyruvate.

### 9.4.1.2 Biosynthesis of vitamins

#### 1. Biosynthesis of niacin:

**Experimental results:** previous work included niacin in all media tested by default (Hussain *et al.* 1991). Our experimental results suggest that this is not an essential vitamin for growth although its absence causes a severe delay.

**Biosynthetic potential and model behaviour:** *de novo* synthesis of niacin (nicotinate) from Asp is not possible in RP62A. However, NAD and nicotinic acid salvage reactions are completed in this strain. Initially, analysis of the model predicted that NAD could be synthesised in the absence of external niacin under aerobic conditions through the O<sub>2</sub>-consuming Asp oxidase reaction (EC 1.4.3.16) in the first step of niacin *de novo* synthesis from Asp. Under anaerobic conditions, exogenous niacin was needed in order to allow NAD production.

**Model refinement:** three reactions catalysed by enzymes involved in *de novo* synthesis of niacin (EC 2.5.1.72, EC 2.4.2.19 and EC 3.5.1.42) had no associated genes present in RP62A, and therefore had been included as a result of automatic gap-filling. For accuracy, these reactions were removed from the model, thus, the inclusion of niacin in the media became unconditionally essential for NAD production.

#### 2. Biosynthesis of thiamine:

**Experimental results:** previous work included thiamine in all media tested by default (Hussain *et al.* 1991). Our experimental results suggest that this vitamin is not essential for growth and its absence has no notable effect.

**Biosynthetic potential and model behaviour:** a set of five enzymes (EC 3.5.99.2, EC 2.7.1.49, EC 2.7.4.7, EC 2.5.1.3 and EC 3.1.3.1) ensure that salvage of thiamine is possible in RP62A. However, *de novo* synthesis from pyridoxal phosphate or 5-aminoimidazole nucleotide (derived from the purine metabolism) is absent in this strain according to BioCyc.

This prosthetic group is not directly involved with the biosynthesis of any other biomass component in the model, therefore, it could be safely omitted in the *in-silico* medium and the cell biomass composition without affecting the results of further analysis of the model.

Since RP62A lacks the ability to perform *de novo* synthesis of niacin and thiamine, the occurrence of delayed growth in their absence could be explained as a consequence of either carry over of small amounts of these vitamins in the inoculum or their uptake from the media after cell-lysis occurring at later time points.

### 3. Biosynthesis of biotin:

**Experimental results:** previous work included biotin in all media tested by default (Hussain *et al.* 1991). Our experimental results suggest that this vitamin is not essential for growth and its absence has no severe effect.

**Biosynthetic potential and model behaviour:** the full pathway for biosynthesis of biotin from pimelate is present in RP62A. The synthesis of pimelate, a seven-carbon dicarboxylic acid, is achieved through the fatty acid synthetic metabolism but the exact reactions that take place and the enzymes involved are yet to be defined (Lin *et al.* 2010), and are not described in the KEGG or MetaCyc databases. Once it is produced, biotin is transformed into biotinyl-5'-AMP and fed back into fatty acid biosynthesis through reactions which again, have not been described.

Since this prosthetic group is not directly involved with the biosynthesis of any other biomass component in the model, it could be safely omitted in the *in-silico* media and the cell biomass composition without affecting the results of further analysis of the model.