Designing obligate consortia with carbon-nitrogen exchange using new-to-nature exchange molecules

Elisabeth Bubna-Litic

2024

Imperial College London

Department of Life Sciences

Thesis Submitted in Partial Satisfaction of the Requirements for the Degree of

Doctor of Philosophy

Abstract

The sustainable provision of nitrogen to agriculture is one of the major challenges facing humanity in the 21st century. A range of biotechnological solutions to replace the highly polluting Haber-Bosch process have been proposed. In this thesis, I introduce an approach using metabolic engineering to design novel nutrient-exchange pathways between crop plants and soil-dwelling nitrogen-fixing microbiota. A consortium of the nitrogen-fixing soil bacterium *Azotobacter vinelandii* with the cyanobacterium *Synechocystis sp.* PCC 6803, a model photoautotrophic organism, is developed both *in vitro* and implemented as genome-scale metabolic model *in silico.* In addition to the total set of compounds naturally present in either consortium partner, rationally designed pathways based on the *Anabaena sp.* PCC 7120 nitrogen-carbon exchange system were evaluated. A high-throughput unbiased approach utilising bioretrosynthesis was also employed to evaluate a broad nutrient-exchange compound design space spanning 1,292,154 compounds including those not native to the consortium partners. Overall, 14,574 unique nutrient-exchange pathways were identified as capable of transferring fixed nitrogen from *A. vinelandii* to *Synechocystis* 6803, and fixed carbon from *Synechocystis* 6803 to *A. vinelandii* in return while supporting growth of both partners, thus enabling the establishment of a nutrient-exchange symbiosis.

Declaration of originality

I, Elisabeth Bubna-Litic, declare that the work presented in this thesis is original work and any collaborations, sources and references are acknowledged appropriately in the text. I further state that no part of my dissertation has been previously submitted for a degree or diploma or other qualification at Imperial College London or any other University or similar institution.

December 20th, 2024

Elisabeth Bubna-Litic

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

I would like to express my deepest gratitude to Prof Patrik Jones for giving me the opportunity to undertake this research. I am also sincerely thankful to my progress review panel members, Dr James Murray and Dr Naomi Nakayama, for their valuable feedback and guidance throughout my PhD, as well as during my undergraduate and master's studies. Special thanks to Prof Geoff Baldwin for his support, particularly during the final months of my PhD. I am also very grateful to Dr Jonathan Muller for his friendship and for keeping our lab running smoothly and teaching me the essential laboratory techniques.

To my family, thank you for your endless support and belief in me, not only during this project but throughout my academic journey. To my friends, thank you for your positivity. Finally, my heartfelt thanks go to my Matyáš – I could not have done it without you.

Contents

Abstract 2

Declaration of originality 3

Copyright declaration 4

Acknowledgements 5

Table of Figures 13

Table of Tables 22

Table of Abbreviations 26

Chapter 1 Introduction 27

1.1 The Nitrogen Crisis 27

1.2 Biotechnological Avenues towards the Fertilisation of Crops 29

1.2.1 The Direct Transfer of Nitrogenase to Plants 29

1.2.2 Endosymbiotic Nitrogen Fixation 39

1.2.3 Rhizosphere enrichment 46

1.3 Designing Obligate Consortia with Carbon-Nitrogen Exchange Using New-to-Nature Exchange Molecules 48

Chapter 2 Materials and Methods 51

2.1 Molecular Biology Methods 51

2.1.1 Amplification of DNA parts 51

2.1.2 Amplification of parts requiring the removal of an internal BsaI sites 53

2.1.3 Gel electrophoresis 54

2.1.4 DNA synthesis 55

2.1.5 DNA part subcloning using CloneJET 55
2.1.6 DNA part subcloning using OE-PCR 56
2.1.7 BASIC Assembly 59
2.1.8 Colony PCR and gel electrophoresis from *E. coli* DH5α 60
2.1.9 Sequencing from *E. coli* DH5α 61
2.2 Cell Culture and Transformation 63
2.2.1 *E. coli* DH5α 63
2.2.2 *Synechocystis sp.* PCC 6803 64
2.2.3 A. vinelandii 65
2.2.4 Coculture of *A. vinelandii* and *Synechocystis* 6803 66
2.2.5 Growth Assays 66
2.2.6 Quantification of Glutamate and Glutamine 68
2.3 Computational Methods 72
2.3.1 Retrieval of Candidate Compounds from Database 73
2.3.2 Conversion of Compound Representations 73
2.3.3 Stoichiometric Modelling 74
2.3.4 Computational Bioretrosynthesis of Candidate Compounds 74
2.3.5 Reaction Reformatting 78
2.3.6 Stoichiometric Balancing of Chemical Equations 79
2.3.7 Compound Clustering 81
Chapter 3 Establishing Cocultures *in vivo* and *in silico* 83

3.1 Introduction 83

3.1.1 Rationale 83

3.1.2 Aims 83

3.1.3 Choice of Chasses 84

3.2 Establishing Cocultures *in vitro* 86

3.2.1 Choosing a Coculture Medium 86

3.2.2 Developing a Method to Discriminate between Growth Rates of Coculture Partners 88

3.3 Creation of Consortium Model *in silico* 98

3.3.1 Genome-scale Metabolic Models and Constraint-Based Modelling 98

3.3.2 Implementation of the Two-Cell Model 100

3.3.3 Media and Other Biologically Relevant Constraints 101

3.4 Model Validation 106

3.5 Summary 109

Chapter 4 Engineering a Nature-Mimicking Nutrient-Exchange System 112

4.1 Introduction 112

4.1.1 Rationale 112

4.1.2 Exchange of Carbon and Nitrogen within the Filamentous Multicellular Cyanobacterium *Anabaena* 7120 112

4.1.3 Aims 114

4.2 Engineering an *Anabaena* 7120-like Nutrient-Exchange System 115

4.2.1 Predicted Growth Rates 117

4.2.2 Cloning and Amplification of Bioparts 119

4.3 Construction of Δ*GOGAT* and Δ*nifL A. vinelandii* strains 126

4.3.1 Cloning a *GOGAT* and *nifL* Knockout Cassette 126

4.3.2 Transformation of *A. vinelandii* 129

4.4 Developing a Method for the Separation and Detection of Glutamate and Glutamine 131

4.5 Summary 137

4.5.1 Transformation of *A. vinelandii* 137

4.5.2 Plasmid Assembly and Transformation of *Synechocystis* 6803 139

4.5.3 Evaluation and Optimisation of the Consortium 140

Chapter 5 Evaluation of the Nutrient-Exchange Potential of Rationally Designed Pathways 142

5.1 Introduction 142

5.1.1 Rationale 142

5.1.2 Aims 142

5.2 Evaluation of Maximum Growth Rates Achievable with Nutrient Exchange Pathways 144

5.2.1 Identification of Candidate Exchange Molecules from a Set of Compounds Native to the Consortium 144

5.3 Growth Rates of Consortia with Rationally Designed Nutrient-Exchange Pathways 147

5.3.1 Deletion of the *A. vinelandii* ammonia reuptake transporter AmtB 147

5.3.2 *A. vinelandii* Δ*amtB* Combined with a Sucrose-Secreting *Synechocystis* 6803 149

5.3.3 Exchange of Glutamine and Glutamate 151

5.3.4 Exchange of Glutamine, Glutamate and Sucrose 153

5.3.5 Glutamine, Glutamate, Sucrose and A-ketoglutarate 155

5.3.6 Glutamine, Glutamate, Sucrose and A-ketoglutarate with *ΔgogatΔgdh A. vinelandii* 157

5.3.7 Summary of Pathways 159

5.4 Summary 162

Chapter 6 Design of Nutrient Exchange Pathways with Bioretrosynthesis 163

6.1 Introduction 163

6.1.1 Rationale 163

6.1.2 Aims 164

6.2 Selection of Candidate Compounds 165

6.3 Bioretrosynthesis-Based Pathway Prediction 167

6.3.1 RetroPathRL: Background on the Bioretrosynthesis Tool Used 167

6.3.2 Implementation and Parameters 172

6.3.3 Pathway Combinations and Scoring 176

6.4 Summary 179

Chapter 7 Implementation and Evaluation of the Nutrient-Exchange Potential of Bioretrosynthetically Identified Pathways 181

7.1 Introduction 181

7.1.1 Rationale 181

7.1.2 Aims 181

7.2 Implementation of Nutrient Exchange Pathways in Metabolic Models 183

7.2.1 Compound Renaming for Compatibility with Cobra Toolbox 3.0 183

7.2.2 Stoichiometric Balancing 183

7.2.3 Implementation 184

7.3 Maximum Biosynthetic Potential 185

7.4 Growth Rates of Consortia with Nutrient-Exchange Pathways 190

7.4.1 Pathways in Media Lacking One Nutrient at a Time 190

7.4.2 Pathways Evaluated in Media Lacking Both Fixed Carbon and Fixed Nitrogen 193

7.5 Summary 197

Chapter 8 Conclusion and Discussion 198

8.1 Establishing a Consortium with Nature-Mimicking Nutrient Exchange 198

8.2 *In silico* Reconstruction of *A. vinelandii*-*Synechocystis* 6803 Consortium 200

8.2.1 Retrosynthesis and a High-Throughput Unbiased Approach to *in silico* Metabolic Engineering 201

8.3 Consortium Robustness to Invasion 208

8.4 Select Pathways for Downstream *In Vitro* Evaluation 209

References 215

Chapter 9 Appendices 227

9.1 Supplementary Data 227

9.1.1 Exemplary Carbon Exchange Pathways 227

9.1.2 Exemplary Nitrogen Exchange Pathways 278

9.1.3 DNA sequences 304

9.1.4 Evaporation assay 308

9.2 Supplementary Methods 309

9.2.1 Evaporation assay 309

9.2.2 Media Preparation 309

9.2.3 Antibiotics Concentrations 312

Table of Figures

Figure 1: The nitrogen cycle. Adapted from Bernhard (2010) 28

Figure 2: The distribution of biological nitrogen fixation across the three domains of living organisms. Clades containing nitrogen fixers are highlighted in purple. No living or extinct eukaryote has been recorded to be capable of directly fixing nitrogen. LUCA denotes last universal common ancestor. Adapted from Boyd & Peters (2013). 30

Figure 3: Structure of MoFe nitrogenase. NifH shown in pink, NifDK in orange and green. Ribbon representation shows cofactors. Figure reproduced from Martin Del Campo et al. (2023). 31

Figure 4: Overview of the maturation of the molybdenum nitrogenase, adapted from Burén et al. (2020). NifM, NifU and NifS are required for the maturation of the Fe protein of nitrogenase (NifH). NifH is in turn required for the maturation of NifDK. Protein subunit colours correspond to Figure 3. 35

Figure 5: A schematic of the distribution of plant diazotrophic symbionts across plant taxa. Both intracellular endosymbionts and extracellular endophytic symbionts are included. 39

Figure 6: Generalised structure of a nod factor. Alternate functional groups at each site are shown in boxes. Adapted from Long (1996). 41

Figure 7: Schematic representation of the layout of primers in OE-PCR for the amplification of bioparts while removing internal BsaI restriction sites. Black rectangle designates genomic region of interest. Blue arrows are outside primers carrying BASIC prefix and suffix sequences. Internal primers are shown in green, overlapping the BsaI recognition site in pink. 54

Figure 8: Schematic representation of primers used in OE-PCR-mediated subcloning. Primers are shown as arrows, where corresponding colours are used to show where on the backbone a section of the primer anneals. 56

Figure 9: Growth curves of A) Synechocystis sp. PCC 6803 and B) Azotobacter vinelandii DJ in BN, BG11, combined media, and water as negative control. Error bars represent the combined standard error of the mean. (N = 3, n = 9) 88

Figure 10: A strong correlation between optical density at 680nm (OD680) and OD730 was found in both A) *Azotobacter vinelandii* DJ ($R^2 \approx 1$) and B) *Synechocystis sp.* PCC 6803 ($R^2 \approx 1$) 89

Figure 11: Calibration curves were prepared for A) *Synechocystis sp.* PCC 6803 at optical density of 730nm (OD730) ($R^2 \approx 0.91$), OD680 ($R^2 \approx 0.93$) and B) *Azotobacter vinelandii* DJ OD730 ($R^2 \approx 0.86$) and OD680 ($R^2 \approx 0.88$). 90

Figure 12: Coculture experiments using *E. coli* in a dialysis pouch in LB medium. Various methods to secure a dialysis pouch were evaluated (twice-folded, corners folded in, double-folded and knotted). Only the bag should contain cells, growth is however apparent in the whole flask, as cells are spilling out. Experiments were carried out and documented by former MRes student Alexandra Cioffi and proposed and supervised by me. 91

Figure 13: Schematic representation of well insert within a well of 6-well plate. The inside / top culture is designated in green, and the outside / bottom culture is lilac.....92

Figure 14: Growth of A) *Azotobacter vinelandii* DJ and B) *Synechocystis sp.* PCC 6803 in 6-well plates with inserts when inoculated inside or outside of a well insert. Stars indicate significance using Student's two-tailed T test, where * = $p > 0.01$, ** = $p > 0.001$, *** = $p > 0.0001$, and **** = $p > 0.00001$. (N = 1, n = 3) 93

Figure 15: Growth curves of *A. vinelandii* wild-type comparing axenic conditions and coculture with *Synechocystis* 6803 wild-type. Cocultures were grown in BNBG11 combination

media perturbations with or without source of carbon as 60 mM sucrose and with or without supplemented nitrogen as 20 mM ammonium acetate. Wild type coculture does not improve growth of *A. vinelandii* without carbon supplementation. Curves fit Gompertz equation (Zwietering et al., 1990). Error bars represent standard deviations for individual experiments (N = 3, n = 9). 95

Figure 16: Growth curves of *Synechocystis* 6803 wild-type comparing axenic conditions and coculture with *A. vinelandii* wild-type. Cocultures were grown in BNBG11 combination media perturbations with or without source of carbon as 60 mM sucrose and with or without supplemented nitrogen as 20 mM ammonium acetate. Wild type coculture does not improve growth of *Synechocystis* 6803 without nitrogen supplementation. Curves fit Gompertz equation (Zwietering et al., 1990). Error bars represent standard deviations for individual experiments (N = 3, n = 9). 96

Figure 17: Density of *A. vinelandii* and *Synechocystis* 6803 cells calculated as hundred million (10^8) cells per millilitre based on Neubauer chamber cell counts. Error bars indicate standard error of the mean. Note the y axis is formatted as $\times 10^8$ (hundred million cells/mL). n = 20 103

Figure 18: A filament of the diazotrophic cyanobacterium *Anabaena* sp. PCC 7120 showing the flow of nutrients between cells in the GS/GOGAT cycle, where GS is glutamine synthase and GOGAT is glutamate synthase. Adapted from Meeks & Elhai (2002). 114

Figure 19: Proposed nutrient exchange mimicking *Anabaena sp.* PCC 7120 using *Synechocystis sp.* PCC 6803 as the photosynthetic partner, and *Azotobacter vinelandii* DJ as the diazotrophic partner, using glutamate (glu) and glutamine (gln) as exchange molecules. A glutamate exporter (yblJ) and sugar exporter (cscB) are introduced into *Synechocystis sp.* PCC 6803 and GOGAT (glutamate synthase) is deleted in *Azotobacter vinelandii* DJ. Water and cofactors are omitted for simplicity. 115

Figure 20: Schematic representation of homologous recombination-mediated gene deletion.....117

Figure 21: Schematic representation of biopart stored using BASIC assembly with methylated linkers (LMP and LMS).....124

Figure 22: Schematic representation of mCherry cassette carrying BBa_J23119 promoter (p BBa_J23119), RBS1 linker, *mCherry* red fluorescent protein gene, neutral linker 3 and rrnB T1 terminator, all flanked by methylated linkers (LMP and LMS) in a pJM03 backbone (pACYC-lac-GFP-GmR).....127

Figure 23: Schematic representation of antibiotic resistance cassette carrying BBa_J23119 promoter (p BBa_J23119), RBS1 linker, kanamycin resistance gene, neutral linker 3 and rrnB T1 terminator, all flanked by methylated linkers (LMP and LMS) in a pJM03 backbone (pACYC-lac-GFP-GmR).....127

Figure 24: Schematic representation of knockout cassette in storage vector, where parts are connected by neutral linkers (L1 and L2), and surrounded by methylated linkers (LMP and LMS). Assembled cassette refers to either the mCherry cassette (Figure 22) or antibiotic resistance cassette (Figure 23) constructed earlier.....128

Figure 25: *A. vinelandii* transformed with GOGAT (Δ*GOGAT::mCherry*) and nifL (Δ*nifL::mCherry*) deletion cassettes carrying mCherry showing a clear pink coloration indicating successful transformation. Plated on antibiotic-free BN agar.....130

Figure 26: No clear peaks are apparent of HPLC chromatographs of glutamate, glutamine or a mixture of the two when detected at 338 nm without derivatisation. All samples at 0.5 g/L concentration. n =3 132

Figure 27: HPLC chromatographs for underivatised glutamate and glutamine at 210 nm show distinct peaks. All samples at 0.5 g/L concentration. n =3 133

Figure 28: HPLC chromatographs for derivatised glutamate and glutamine at 210 nm show distinct peaks, which can however not be used to discriminate between the two amino acids in mixture due to peak overlap. All samples at 0.5 g/L concentration. n =3 134

Figure 29: HPLC chromatographs for derivatised glutamate and glutamine at 338 nm show distinct peaks that can be used to distinguish between the amino acids in mixture and quantify them separately. All samples at 0.5 g/L concentration. n =3 135

Figure 30: HPLC calibration curve with OPA derivatisation detected at 338 nm for A) glutamate ($R^2 = 0.88$) and B) glutamine ($R^2 = 0.88$), where A is integrated peak area (arbitrary units), and c is concentration (g/L). 136

Figure 31: Maximum predicted growth rate of *Synechocystis* 6803 in consortium in nitrogen-free, carbon-free media for each combination of carbon-donor and nitrogen-donor pathways predicted as maximum biomass objective function flux (gDW/h). Compounds are labelled using BiGG identifier for brevity (for compound names, see Supplementary File S3). 145

Figure 32: Maximum predicted growth rate of *A. vinelandii* in consortium in nitrogen-free, carbon-free media for each combination of carbon-donor and nitrogen-donor pathways predicted as maximum biomass objective function flux (gDW/h). Compounds are labelled using BiGG identifier for brevity (for compound names, see Supplementary File S3) 146

Figure 33: Schematic representation of a nutrient exchange pathway for ammonia involving the deletion of the AmtB ammonia reuptake transporter in *A. vinelandii* and transport of ammonia between cells.....148

Figure 34: Schematic representation of a nutrient exchange pathway for ammonia involving the deletion of the AmtB ammonia reuptake transporter in *A. vinelandii,* the introduction of a sucrose transporter in *Synechocystis* 6803 and transport of ammonia and sucrose between cells. 150

Figure 35: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine and glutamate between *A. vinelandii* and *Synechocystis* 6803.....152

Figure 36: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine, glutamate and sucrose between *A. vinelandii* and *Synechocystis* 6803. 154

Figure 37: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between *A. vinelandii* and *Synechocystis* 6803.....156

Figure 38: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between *A. vinelandii* and *Synechocystis* 6803, in addition to a deletion of GDH and GOGAT in *A. vinelandii*.....158

Figure 39: Heatmap of maximum biomass objective function flux for *Synechocystis* 6803 in consortium using different nutrient exchange pathways. 160

Figure 40: Heatmap of maximum biomass objective function flux for *A. vinelandii* in consortium using different nutrient exchange pathways. 161

Figure 41: Schematic representation of the phases of the Monte Carlo Tree Search Reinforcement learning method. Adapted from Koch (2019) 171

Figure 42: Flow diagram representing the selection process of carbon-containing compounds from the MetaNetX database to those which could be synthesised and broken down in the respective chasses. 174

Figure 43: Flow diagram representing the selection process of nitrogen-containing compounds from the MetaNetX database to those which could be synthesised and broken down in the respective chasses. 175

Figure 44: Heatmap representing the pathway lengths and biochemical scores of the designed nitrogen-carrying pathways. 177

Figure 45: Heatmap representing the pathway lengths and biochemical scores of the designed carbon-carrying pathways.....178

Figure 46: Histograms showing a distribution of maximum pathway flux (mmol/gDWh) through final reaction within each biosynthetic pathway in rich media in A) *Synechocystis* 6803 constrained to reflect photoautotrophic conditions, B) *A. vinelandii* constrained to reflect diazotrophic conditions, C) *Synechocystis* 6803 with free hydrogenase flux, and D) *A. vinelandii* unconstrained for ATP maintenance.....187

Figure 47: Correlation of predicted maximum pathway flux (mmol/gDWh) as proxy for maximum compound yield for compounds in *Synechocystis* 6803 in rich media with photoautotrophy-specific constraints and *Synechocystis* 6803 in rich media unconstrained for photoautotrophy. Each marker represents a single pathway. Linear regression calculated using only pathways for which maximum predicted flux is positive in both constrained and unconstrained conditions (green markers) shows a correlation coefficient $R^2 = 0.998$.....188

Figure 48: Correlation of predicted maximum pathway flux (mmol/gDWh) as proxy for maximum compound yield for compounds in *A. vinelandii* in rich media with diazotrophy-specific constraints and *A. vinelandii* in rich media unconstrained for photoautotrophy. Each marker represents a single pathway. Linear regression calculated using only pathways for which maximum predicted flux is positive in both constrained and unconstrained conditions (violet markers) shows a correlation coefficient $R^2 = 0.999$. 189

Figure 49: Histograms representing distribution of maximum growth rate (calculated as biomass objective function flux gDW/h) in A) *Synechocystis* 6803 in consortium in nitrogen-free media and B) *A. vinelandii* in consortium in carbon free media, C) *A. vinelandii* in

consortium in nitrogen-free media, and D) *Synechocystis* 6803 in consortium in carbon-free media.....192

Figure 50: Distribution of the maximum biomass flux for A, C) *Synechocystis* 6803 in nitrogen-deplete media and B, D) *A. vinelandii* in carbon-deplete media in consortium with nutrient-exchange pathways for the respective essential nutrient. Figures C and D represent zoomed-in version of A and B for visual clarity. Box plots overlaid over individual datapoints extend from the first quartile to the third quartile of the data, with a line on the median. Whiskers extend to the farthest point within 1.5x the interquartile range from the box. 194

Figure 51: A heatmap representing the growth rates of *Synechocystis* 6803 in carbon-free, nitrogen-free media representing a combination of 36 nitrogen-carrying and 447 carbon-carrying pathways in the consortium. Only pathways with a positive maximum growth rate for both species are represented. Darker shades represent higher growth rates. Colour scale is trimmed at an upper limit of 1 for higher visual clarity; for the full range of growth rates, refer to Figure 50A. 195

Figure 52: A heatmap representing the growth rates of *A. vinelandii* in carbon-free, nitrogen-free media representing a combination of 36 nitrogen-carrying and 447 carbon-carrying pathways in the consortium. Only pathways with a positive maximum growth rate for both species are represented. Darker shades represent higher growth rates. Colour scale is trimmed at an upper limit of 1 for higher visual clarity; for the full range of growth rates, refer to Figure 50B. 196

Figure 53: UMAP/hdbscan clustering analysis of A-C) nitrogen carrying compounds and D-F) carbon-carrying compounds in media colored by A, D) cluster, B, E) maximum *Synechocystis* 6803 biomass flux (gDW/h) and C, F) maximum *A. vinelandii* biomass flux (gDW/h). 203

Figure 54: Comparison of the distribution of maximum growth rates predicted for *Synechocystis* 6803 (left) and *A. vinelandii* (right) in consortium in media lacking both nitrogen and carbon between pathways designed using bioretrosynthesis (new) and derived from the set of nitrogen and carbon containing compounds already present in the wild-type (native). The top graphs represent a zoomed-out version of graphs containing all datapoints including pathways showing zero growth, bottom two graphs are presented for visual clarity. 205

Table of Tables

Table 1: Abbreviations of gene clusters involved in nitrogen fixation and a short description. Summarised from Martin Del Campo et al. (2023). 32

Table 2: Reaction mixture for amplification of bioparts from genomic DNA 51

Table 3: Cycling protocol for the amplification of bioparts from genomic DNA for cloning..... 52

Table 4: Annealing temperatures for the amplification of bioparts from genomic DNA for cloning..... 52

Table 5: Reaction mixture for the 2nd amplification of bioparts requiring the removal of an internal BsaI site from genomic DNA 54

Table 6: Modified reaction mixture for blunt end cloning using the CloneJET kit 55

Table 7: Reaction mixture for DNA part subcloning using OE-PCR, stage 1 57

Table 8: Cycling protocol for DNA part subcloning using OE-PCR, stage 1..... 58

Table 9: Reaction mixture for DNA part subcloning using OE-PCR, stage 2 58

Table 10: Cycling protocol for DNA part subcloning using OE-PCR, stage 2..... 59

Table 11: BASIC Digestion/Ligation Reaction Setup 59

Table 12: BASIC Assembly reaction 60

Table 13: Reaction mixture for colony PCR..... 61

Table 14: Cycling protocol for colony PCR 61

Table 15: Autosampler program for online OPA derivatisation..... 69

Table 16: HPLC setup details 70

Table 17: HPLC flow protocol..... 71

Table 18: Python packages used in this study..... 72

Table 19: Cobra Toolbox functions used 74

Table 20: Parameter Choice for RetroPathRL. Adapted from RetroPathRL GitHub documentation (Koch, 2019). 76

Table 21: Genome and metabolic model accession numbers for strains used in this study 85

Table 22: Comparison of carbon and nitrogen sources in media for *A. vinelandii* and *Synechocystis* 6803 87

Table 23: Recipe for combination medium for the coculture of *A. vinelandii* and *Synechocystis* 6803. The addition of Bacto Agar is necessary to prepare solid culture plates. 87

Table 24: Maximum growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 in BNBG11 media perturbations, as calculated from Gompertz fit (see Figure 15, Figure 16). Empty cells indicate no growth. 97

Table 25: Biomass functions available in the joint consortium model 101

Table 26: Concentrations (mM) of compounds and ions in media as implemented in setMediumConstrains(). 102

Table 27: Cell concentration and weight as input into the model based on Figure 17. 104

Table 28: Flux bounds for *Azotobacter vinelandii* and *Synechocystis* 6803 set from literature. (mmol/gDWh) 105

Table 29: Expected growth profiles in different media perturbations 106

Table 30: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 in BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells for visual clarity. Compare with Table 24 for *in vitro* data. 108

Table 31: Genetic modifications of *A. vinelandii* and *Synechocystis* 6803 to induce a nature-mimicking exchange 116

Table 32: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium engineered to mimic the nutrient exchange in *Anabaena* 7120 per Figure 19 and Table 31. Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison. 119

Table 33: DNA bioparts ordered for synthesis for nature-mimicking nutrient exchange. All sequences are listed in Supplementary table 1..... 121

Table 34: Plasmid backbones used in this work 125

Table 35: List of DNA parts and subcloning strategies 125

Table 36: *A. vinelandii* constructs under construction in the present work 139

Table 37: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with deletion of *A. vinelandii amtB* per Figure 33..... 149

Table 38: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with deletion of *A. vinelandii amtB* and export of sucrose from *Synechocystis* 6803 per Figure 34..... 151

Table 39: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine and glutamate between cells per Figure 35. 153

Table 40: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine, glutamate and sucrose between cells per Figure 36. 155

Table 41: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between cells per Figure 37. 157

Table 42: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between cells per Figure 38. 159

Table 43: Subsets of compounds retrieved from MetaNetX database used in this analysis.....166

Table 44: Numbers of compounds and pathways with a positive non-zero maximum pathway flux when optimised for.....186

Table 45: Mean, lower quartile (Q1) and upper quartile (Q3) of maximum growth rates (gDW/h) predicted for *Synechocystis* 6803 and *A. vinelandii* in consortium in media lacking both nitrogen and carbon between pathways designed using bioretrosynthesis (new) and derived from the set of nitrogen and carbon containing compounds already present in the wild-type (native). 204

Table 46: Top filtered pathway combinations for downstream *in vitro* evaluation. Carbon and nitrogen exchange pathways are identified using a pathway ID, for a full graphical representation of pathways, see Supplementary data, Sections 8.5.1 and 8.5.2. 211

Table of Abbreviations

Abbreviation	Explanation
ARA	Acetylene Reduction Assay
BASIC	Bioparts Assembly for Idempotent Cloning
BNF	Biological Nitrogen Fixation
DW	Dry weight
FBA	Flux Balance Analysis
FLNF	Free-living Nitrogen Fixer
GEM	Genome-scale Metabolic Model
gDNA	Genomic DNA
OE-PCR	Overlap Extension PCR
PCR	Polymerase Chain Reaction
PGPRs	Plant Growth-Promoting Rhizobacteria
SBML	Systems Biology Markup Language

Chapter 1
Introduction

1.1 The Nitrogen Crisis

The availability of fixed nitrogen is a limiting factor to plant growth and thus also agricultural productivity. Biological nitrogen fixation (BNF), the process in which atmospheric nitrogen gas (N_2) is reduced to form bioavailable ammonia, is, however, limited to prokaryotes (Dos Santos et al., 2012). Food crops rely on nitrogen in the form ammonia supplied by diazotrophic bacteria, nitrates and nitrites produced from ammonia by nitrifying bacteria, and amino acids (Figure 1). Since the agricultural revolution, agricultural history, humans have increased yields of food crops using supplemented nitrogen in the form of fertiliser such as manure, green manure, intercropping and crop rotation.

Today, synthetic fertiliser is largely produced using the Haber-Bosch process, which employs high temperature and pressure with inorganic catalysis to reduce atmospheric (N_2) to ammonia (Haber, 1920). The Haber-Bosch process has become instrumental in providing nearly half of the calories consumed by humanity over the course of the 20th and early 21st century (Erisman et al., 2008), and with growing population and wealth in developing countries, humanity's demand for fixed nitrogen will continue to grow (Galloway et al., 1995; Koning et al., 2008). Without human intervention, the sources of bioavailable nitrogen are limited to the nitrogen fixed by nitrogen-fixing bacteria, nitrogen recycled through the nitrogen cycle by nitrifying and denitrifying bacteria, and an insignificant amount of nitrogen generated during lightning strikes.

The development of alternatives to synthetic nitrogen fertilisers is necessary for transitioning away from a fossil fuel-based agriculture, as their production is expensive, energetically demanding, and detrimental to the environment, contributing to the release of

ammonia and nitrogen oxides to the atmosphere. Most of the energy used in nitrogenous fertiliser synthesis throughout Europe also comes from natural gas (Woods et al., 2010), tying food prices to the cost of fossil fuels, and adding vulnerability to agricultural supply chains in times of natural gas supply uncertainty. At the same time, non-specific fertilisation of fields results in weed growth and therefore an increased requirement for herbicides. Excessive use of fertiliser results in run-off from fields which leads to eutrophication, soil acidification, and a loss of biodiversity (Trinh et al., 2015). These challenges call for the development of a novel nitrogen fertilisation strategy which avoids both being energy demanding and the need for excessive use of fertiliser on fields.

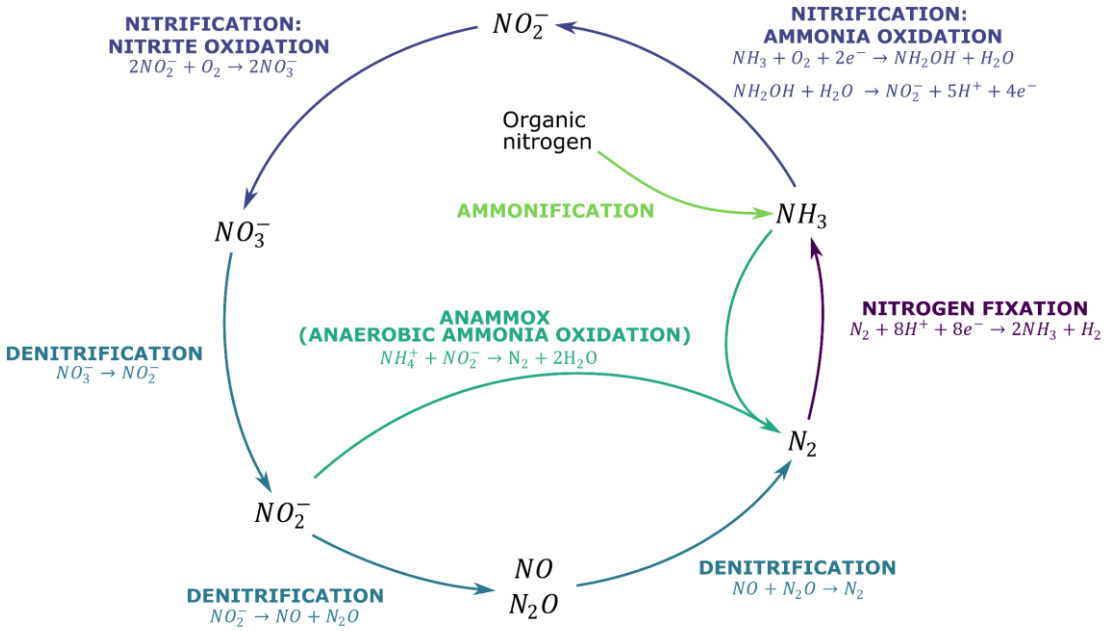


Figure 1: The nitrogen cycle. Adapted from Bernhard (2010)

1.2 Biotechnological Avenues towards the Fertilisation of Crops

Three main avenues for more effective delivery of bioavailable nitrogen to crops exist in today's biotechnology: 1) the direct transfer of nitrogenase to plants, 2) introduction of diazotrophic symbionts into the plants – either as intercellular endosymbionts, or extracellular endophytes, and 3) the enrichment of the rhizosphere with free-living nitrogen fixers (Beatty & Good, 2011; Charpentier & Oldroyd, 2010; Curatti & Rubio, 2014; Mus et al., 2016).

1.2.1 The Direct Transfer of Nitrogenase to Plants

1.2.1.1 The Nitrogenase Enzyme

Biological nitrogen fixation (BNF) is a complex biological process under stringent genetic regulation. The core reaction of BNF – the reduction of atmospheric nitrogen (N_2) to ammonia (NH_4^+) – is catalysed by nitrogenase, a complex metalloenzyme with a complicated bioassembly route that is yet to be fully elucidated (Martin Del Campo et al., 2023). BNF is limited to prokaryotes, distributed across a diversity of bacteria and archaea (Figure 2) (Dixon & Kahn, 2004; Dos Santos et al., 2012). Plants instead take up nitrogen from the soil primarily as nitrate (NO_3^-) and ammonia (NH_4^+) using a set of transporters within the root system. Several amino acid transporters within the root system also enable direct uptake of organic sources of amino acids. The overall nitrogen metabolism of plants has recently been reviewed by Zayed et al. (2023)

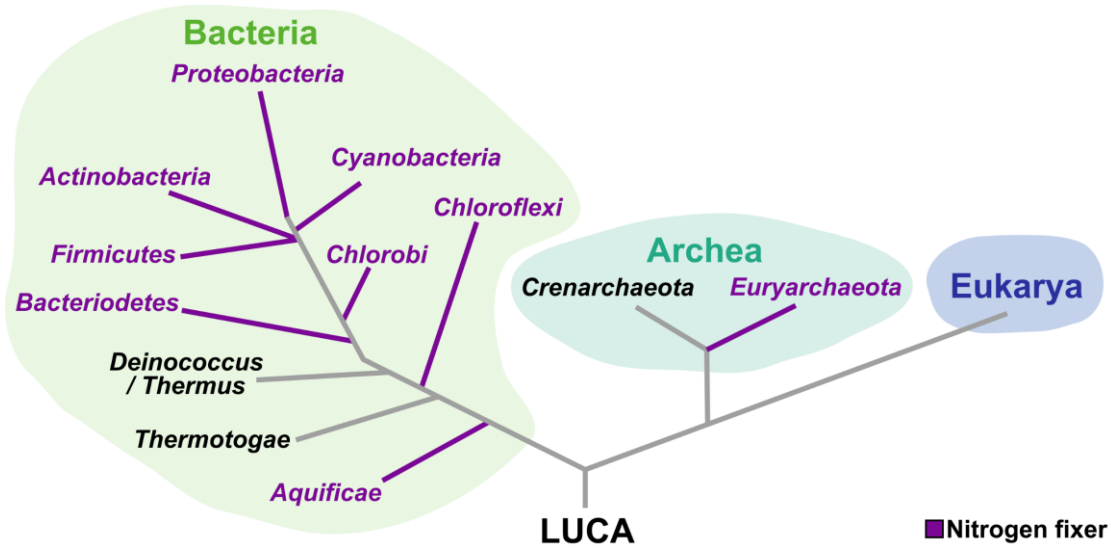


Figure 2: The distribution of biological nitrogen fixation across the three domains of living organisms. Clades containing nitrogen fixers are highlighted in purple. No living or extinct eukaryote has been recorded to be capable of directly fixing nitrogen. LUCA denotes last universal common ancestor. Adapted from Boyd & Peters (2013).

All nitrogen-fixing bacteria (diazotrophs) have a highly conserved molybdenum-iron nitrogenase (often referred to as MoFe); other vanadium-iron (VFe) or iron-iron (FeFe) nitrogenase are expressed by some diazotrophs under nutrient-limited conditions. Nitrogenase was first isolated from *Azotobacter vinelandii* by Bulen & Lecomte (1966). All types of nitrogenases show a high degree of sequence and structure conservation across the three different types (Eady, 1996; Zehr et al., 2003). The MoFe nitrogenase is a dimer consisting of two proteins: NifDK and NifH. NifDK is an $\alpha_2\beta_2$ heterotetramer carrying one iron-molybdenum cofactor (FeMo-co) [7Fe-9S-C-Mo-*R* homocitrate] in the active site of each α subunit, and one P-cluster [8Fe-7S] at the α/β interface. NifH is a homodimer with a MgATP binding site at the dimer interface and an [4Fe-4S] cluster at the subunit interface (Figure 3) (Rubio & Ludden, 2008). Due to these exposed iron-sulphur clusters, nitrogenases are universally (Mackellar et al., 2016) extremely oxygen sensitive enzymes (Gallon, 1981).

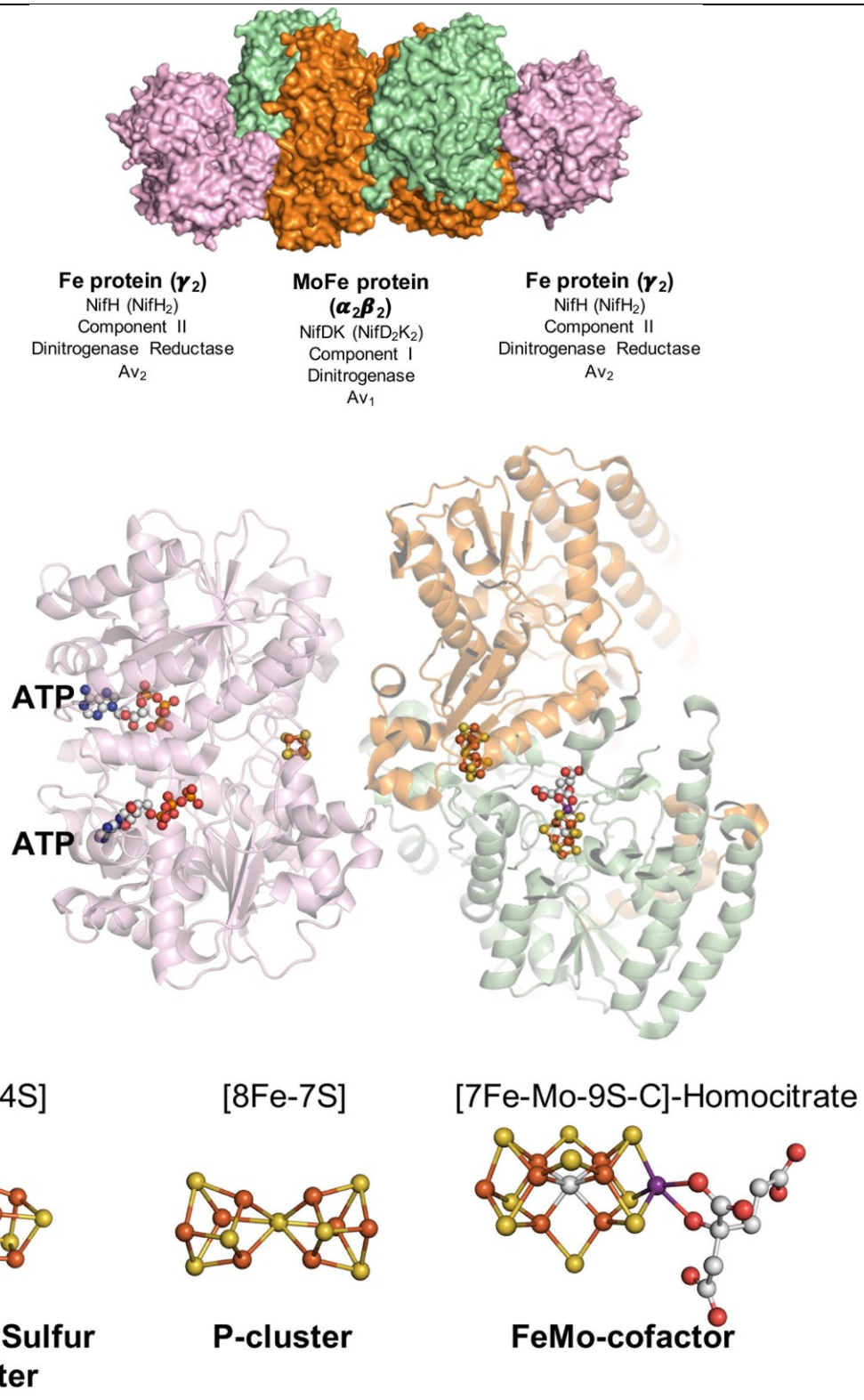


Figure 3: Structure of MoFe nitrogenase. NifH shown in pink, NifDK in orange and green. Ribbon representation shows cofactors. Figure reproduced from Martin Del Campo et al. (2023).

While the nitrogenase protein itself is highly conserved, considerable diversity exists in the biosynthesis and regulation of nitrogenase among different diazotrophs. Genes encoding MoFe nitrogenase subunits or directly related to its bioassembly form the *nif* gene cluster, were first described in *Klebsiella pneumoniae*, in which this cluster forms a contiguous section of the chromosome, by Arnold et al. (1988). Several other gene clusters are involved in the native functioning of nitrogen fixation in diazotrophs. For example, in *A. vinelandii*, 103 genes are directly involved in the activity of its three nitrogenase isoforms and the regulation of nitrogen fixation. Additionally, these genes are not in near physical proximity on the chromosome (Martin Del Campo et al., 2023). Notably, the *vnf* cluster is essential for the VFe nitrogenase and *anf* for the FeFe nitrogenase; other gene clusters necessary for bioassembly, protein activity regulation, regulation of gene expression, or cofactor transport and storage are summarised in Table 1.

Table 1: Abbreviations of gene clusters involved in nitrogen fixation and a short description. Summarised from Martin Del Campo et al. (2023).

Abbreviation	**Definition**	**Description**
anf	alternative nitrogen fixation	Genes associated with the iron-iron (FeFe) nitrogenase (originally known as the "second alternative nitrogenase")
fix	Nitrogen fixation	Other genes involved in nitrogen fixation; some are involved in respiratory protection and electron transfer
mod	molybdenum	Molybdenum transport and acquisition for MoFe nitrogenase

mos	molybdenum storage	Genes associated with molybdenum storage
naf	nitrogenase associated factor	Other genes associated with the *nif* cluster
nfa	nitrogen fixation associated	Other genes associated with nitrogen fixation
nif	nitrogen fixation	Genes associated with molybdenum-iron (MoFe) nitrogenase
rnf	*Rhodobacter* nitrogen fixation	Cluster of genes co-located on the genome with one of the *nif* gene clusters (note: originally isolated from *Rhodobacter capsulatus*, but also exist in others)
vnf	vanadium nitrogen fixation	Genes associated with vanadium-iron (VFe) nitrogenase
vod	*vnf* (paralogous to *mod*)	Genes associated with the *vnf* genomic region

Nitrogen fixation is a highly energy demanding reaction: under optimal conditions, 16 molecules of ATP are necessary for the reduction of one molecule of N_2 (Dixon & Kahn, 2004); metabolic readjustments in response to oxygen stress can drive this as high as 54 ATP per N_2 (Stam et al., 1984). Nitrogenase is also a relatively slow enzyme with a turnover time of ~5 s^{-1} (Dixon & Kahn, 2004). For comparison, the turnover rate of the notoriously slow RuBisCO is ~3 s^{-1} (Ma et al., 2014). To synthesise sufficient levels of bioavailable nitrogen, a large amount of enzyme is therefore required – in diazotrophic conditions, nitrogenase takes up as much as

20% of the total protein content of the cell (Dixon & Kahn, 2004). See Martin Del Campo et al. (2023) for a recent in-depth review of the nitrogenases, their mechanism of action, bioassembly and regulatory cascades.

1.2.1.2 The Expression of Nitrogenase in Plants

Several main challenges must be overcome to transfer nitrogenase into a plant chassis: 1) the complex, yet to be fully understood bioassembly of nitrogenase; 2) the oxygen sensitivity of the enzyme; 3) the stringent regulation natural biological nitrogen fixation is under; and 4) the energy and cofactor requirements of the catalysis itself.

To transfer nitrogenase to plant cells, *nifH*, *nifD* and *nifK*, the genes encoding the Fe protein (NifH) and the MoFe protein (NifDK) of the molybdenum nitrogenase, are the bare minimum set of genes that must be expressed. The biosynthesis of the nitrogenase enzyme and the required cofactors is a multi-step, tightly regulated process involving an extensive set of additional proteins *in vivo* especially due to the complicated cofactors necessary for electron transfer during the nitrogenase catalytic cycle. A brief overview of the bioassembly of the molybdenum nitrogenase is presented in Figure 4. Two times three sets of iron-sulphur clusters are required for molybdenum nitrogenase activity: [7Fe-9S-C-Mo-*R* homocitrate] (FeMo-Co), [8Fe-7S] (P-cluster) and [4Fe-4S] (Figure 3). The biosynthesis of nitrogenase cofactors has recently been reviewed by Burén et al. (2020).

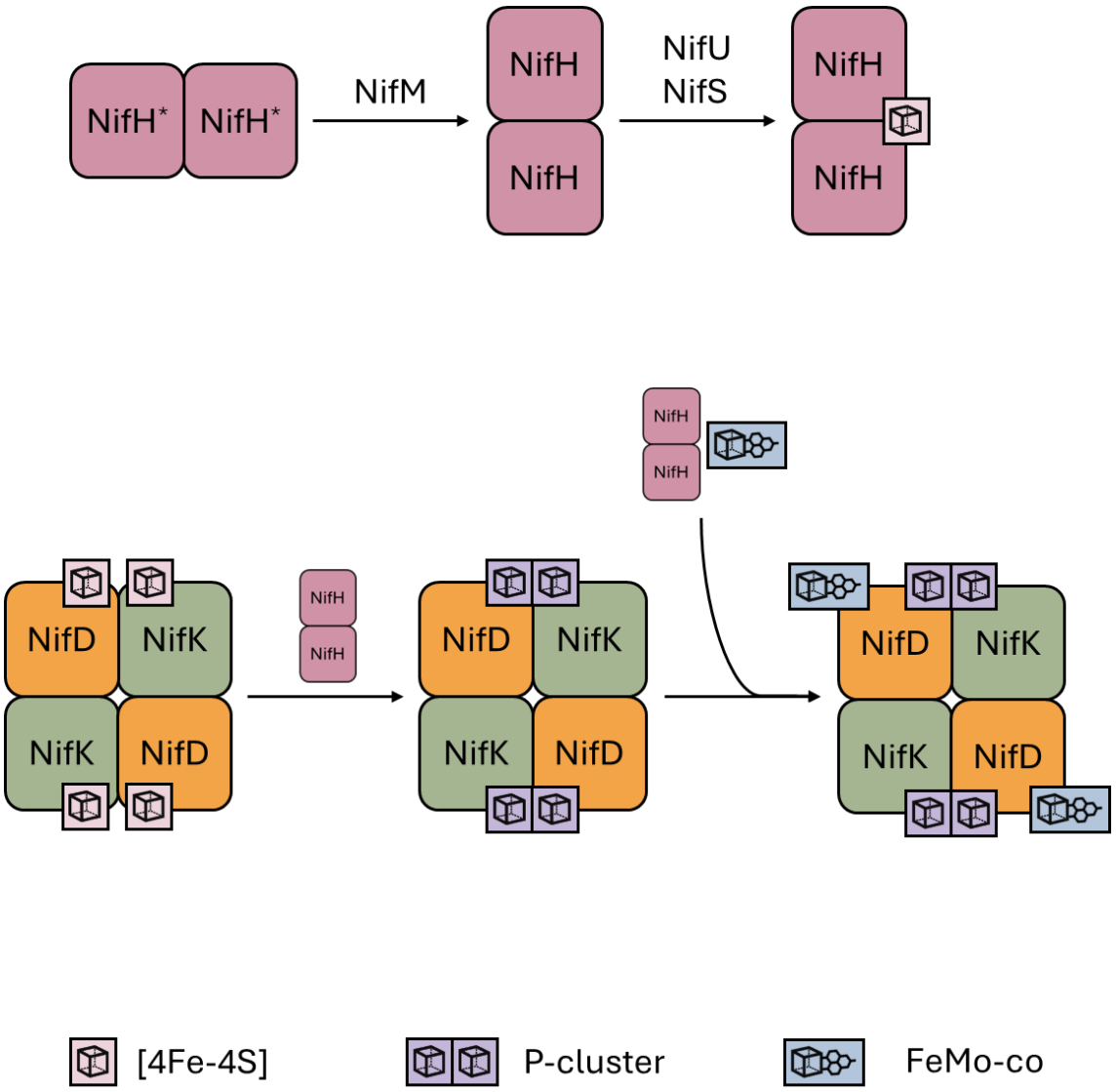


Figure 4: Overview of the maturation of the molybdenum nitrogenase, adapted from Burén et al. (2020). NifM, NifU and NifS are required for the maturation of the Fe protein of nitrogenase (NifH). NifH is in turn required for the maturation of NifDK. Protein subunit colours correspond to Figure 3.

Most efforts to transfer nitrogenase to plants have involved the molybdenum nitrogenase, focusing specifically on the Fe protein (NifH), whose biosynthesis is simpler and better characterised than that of the MoFe protein (NifDK). Extensive work has also been undertaken to determine a subset of additional *nif* genes essential for nitrogenase bioassembly in non-native systems.

The first transfer of nitrogen fixation into a non-diazotroph, namely from *Klebsiella pneumoniae* to *E. coli*, was achieved by Dixon & Postgate (1972) through intergenic mating. Nitrogenase activity under anaerobic conditions was confirmed using acetylene reduction assay (ARA), which is the gold standard for evaluating nitrogenase activity. Curatti et al. (2007) have demonstrated the complete synthesis of the iron-molybdenum cofactor (FeMo-co) of the molybdenum nitrogenase in a cell-free system anaerobic system using only purified NifB, NifEN and NifH proteins and Fe^{2+}, S^{2-}, MoO_4^{2-}, and *R*-homocitrate, *S*-adenosyl methionine, and Mg-ATP. Their work also suggests that NafY and NifX may not be essential to the apoenzyme assembly in anaerobic conditions.

Localisation of nitrogenase expression to plastids or to the mitochondrion to protect nitrogenase from oxygen has originally been proposed by Merrick & Dixon (1984), possibly coupled with temporal separation of nitrogen and carbon fixation if localised to the chloroplast, as is the case in some non-heterocystous diazotrophic cyanobacteria (Dixon et al., 1997). Localisation of nitrogenase to the mitochondrion allows for a similar respiratory oxygen protection mechanism as that employed by *A. vinelandii*. López-Torrejón et al. (2016) were able to express active NifH protein in the mitochondria of aerobically grown yeast *Saccharomyces cerevisiae* when co-expressed with the NifM maturase. Notably, this work demonstrates the ability of NifH to recruit endogenous mitochondrial Fe-S clusters. In contrast with this, cytosolic expression of NifH, which would only remain functional under anaerobic conditions, was only possible when co-expressed with NifU and NifS, in addition to NifM. Burén et al. (2017) successfully expressed nine *nif* genes (*nifHDKUSMBEN*) in yeast mitochondria and confirmed mitochondrial of P-cluster-containing apo-NifDK; the protein could however not be activated by FeMo-co *in vitro*, indicating that further fine-tuning of expression levels and timings, or the expression additional biosynthetic proteins is necessary to assembly functional NifDK in mitochondria.

Active NifH protein was successfully expressed in the chloroplasts of tobacco (*Nicotiana benthamiana*) plants. Eseverri et al. (2020) have demonstrated accumulation of soluble NifH, NifM, NifU and NifS in the chloroplast stroma of tobacco mesophyll cells, demonstrating NifH activity when protein isolated during the dark period (for temporal separation from oxygenic photosynthesis) was incubated *in vitro* with *A. vinelandii* MoFe (NifDK) protein, showing ARA activity. *In vitro* reconstitution with iron-sulphur clusters was shown to significantly improve ARA activity relative to incubation with *A. vinelandii* NifDK only, suggesting a low, but non-zero occupancy of iron-sulphur clusters in chloroplast-matured NifH. More recently, Baysal et al. (2022) have also reported expression of NifH in rice (*Oryza sativa*) mitochondria when co-expressed with NifM. The NifH protein showed limited catalytic activity due to the poor occupancy of iron-sulphur clusters; this could be rescued *in vitro* by the iron-sulphur donor NifU. Iron-sulphur cluster maturation indeed appears to be the bottleneck in achieving active NifH expression in plant cells (Jiang et al., 2021).

Recently, additional optimisation approaches have been used to overcome challenges in transferring nitrogen-fixation to plants. Notably, Smanski et al. (2014) have applied a highly parallelised design-build-test-learn process to refactor the nitrogen fixation (*nif*) gene cluster from *Klebsiella oxytoca*. The native cluster spans 25kb with 20 genes in 7 operons. The diversity of architecture of *nif* clusters across a variety of nitrogen fixing organisms has played a role in their choice to pursue a combinatorial, as there was little prior evidence to the advantages of one design over another. Interestingly, their work did not suggest underlying rules governing what makes one architecture “better” than another (Smanski et al., 2014), providing further support for combinatorial approaches.

To decrease the number of genes requiring transformation and to achieve balanced expression levels, Yang et al. (2018) have devised a “fuse-and-cleave” strategy. 14 genes of the *K. oxytoca nif* cluster were reconstituted as 5 giant genes to be post-translationally cleaved

using Tobacco Etch Virus protease (TEVp), grouping genes by stoichiometry and expression levels. This strategy was successful in enabling diazotrophic growth in *E. coli*. The strategy was later extended to enable mitochondrial expression of nitrogenase in baker's yeast *Saccharomyces cerevisiae* by replacing TEVp cleavage sites with peptide sequences susceptible to cleavage by endogenous mitochondrial processing peptidase, although no nitrogenase activity was detected in the yeast (Yang et al., 2023).

In spite of several promising developments in the work to transfer nitrogen fixation to plants, to date, neither of the two proteins that make up the molybdenum nitrogenase has successfully been assembled and populated with cofactors in a eukaryotic system. Current research suggests that the biggest hurdle to overcome is the complex bioassembly of nitrogenase and its associated cofactors. The balance of protein expression, protein stability and the provision of sufficient precursors and energy molecules are likely to become additional challenges if NifDK and NifH expression in eukaryotic systems ever becomes viable.

1.2.2 Endosymbiotic Nitrogen Fixation

While BNF is limited to prokaryotes, multiple clades of plants have evolved to form tight symbiotic relationships with endophytic diazotrophic bacteria (Figure 5). In this context, endosymbiosis is used to refer to intercellular symbioses found in the legume-rhizobia symbiosis, as well as extracellular endophytic symbioses. Endophytes are bacteria which colonise plant tissue internally and cannot be removed by surface sterilisation. Symbioses and associations outside the plant body (in the soil surrounding plant root systems) are discussed in Section 1.2.3. A major advantage endosymbiotic bacteria have over free-living ones is the protection from environmental competition.

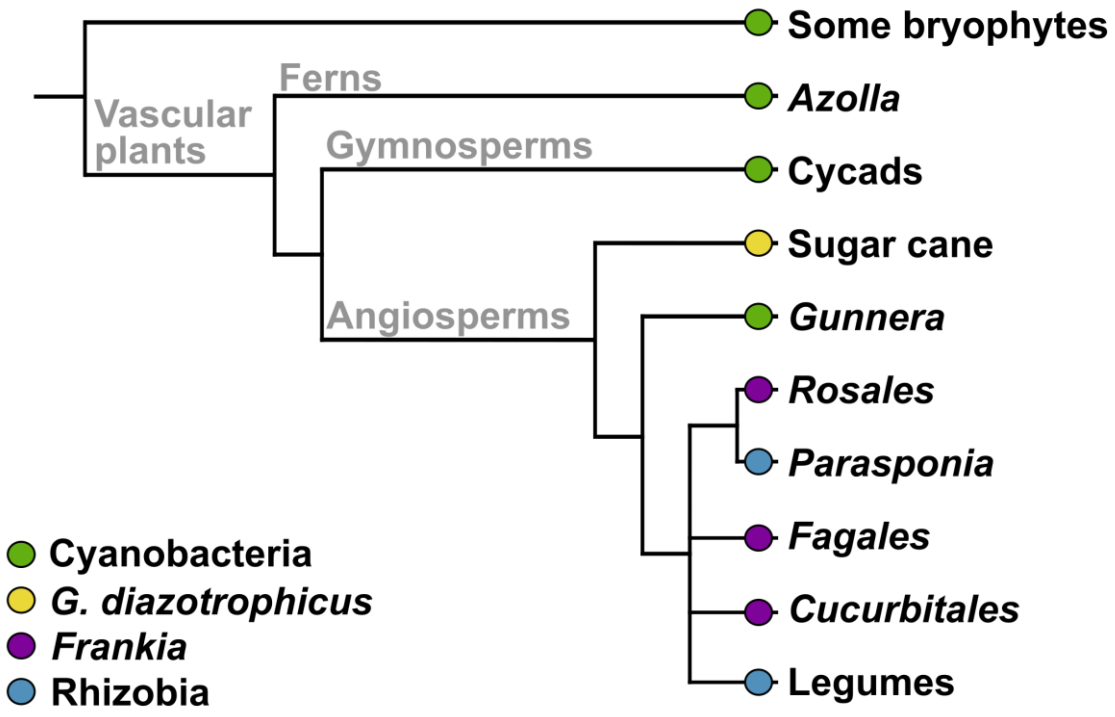


Figure 5: A schematic of the distribution of plant diazotrophic symbionts across plant taxa. Both intracellular endosymbionts and extracellular endophytic symbionts are included.

1.2.2.1 The Symbiosis of Legumes with Rhizobia

The most widely studied of these relationships is the obligate association of rhizobia, a paraphyletic group of α- and β-proteobacteria, with their legume hosts. Outside of legumes, rhizobia also associate with trees of the dicot genus *Parasponia* (Behm et al., 2014).

The infection of legumes by rhizobia involves reciprocal signalling and tight regulation of developmental reprogramming. The developmental processes and rules of engagement required for the establishment of nodule-based symbiosis was reviewed by Oldroyd & Downie (2008) and Oldroyd et al. (2011). In summary, root hair cells of legumes curl around infecting rhizobia, entrapping them. Entrapped bacteria subsequently grow, forming infection threads – plant-derived invaginations crossing cell boundaries, which facilitate the invasion of rhizobia into cells of the root cortex. Depending on the type of infection thread, plant cytoskeletal remodelling can be necessary to facilitate the transversal of cells. At the same time, cortical cells divide, forming nodule primordia. Infection threads direct rhizobia to these, and the rhizobia are released into the inner cells of the nodule. Rhizobia are then encapsulated within the plant membrane as they get taken up by an endocytic mechanism. Finally, endosymbiosis-derived membranes and rhizobia differentiate to form symbiosomes, organelle-like structures hosting bacteroids – terminally differentiated bacteria.

The developmental reprogramming associated with nodulation is induced by signal lipo-chitooligosaccharides termed Nod factors (Figure 6) released by invading rhizobia, which are specific to each legume-rhizobium pairing and crucial to determining the host range specificity in legume-rhizobia symbioses (Walker et al., 2020). Additionally, attachment of rhizobia to the root epidermis is facilitated by the interaction of plant lectins (highly specific carbohydrate-binding proteins) with specific polysaccharides on the bacterial surface; alternative lectins have been shown to change nodulation specificity between different legume and rhizobia species (van Rhijn et al., 1998).

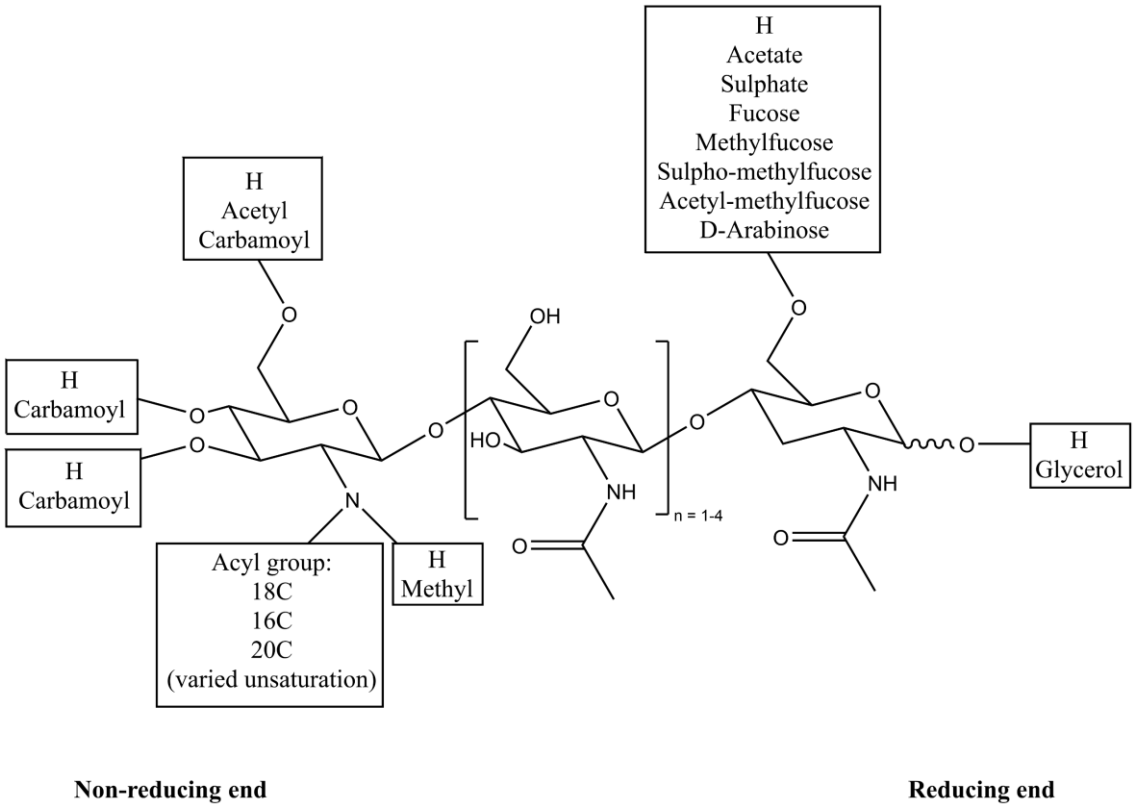


Figure 6: Generalised structure of a nod factor. Alternate functional groups at each site are shown in boxes. Adapted from Long (1996).

Non-cooperative rhizobia are sanctioned by the host plant. Kiers et al. (2003) have used a nitrogen-deplete atmosphere to observe that when *Bradyrhizobium japonicum* in soybean (*Glycine max*) cannot fix nitrogen, and thereby "cheats" in its symbiosis with the plant, its reproductive success significantly decreases. In the absence of plant-imposed sanctions, bacteroid mutants that do not waste resources on providing fixed nitrogen to the plant partner and only benefit from the nutrients provided by the plant would be expected to grow better than "honest" bacteroids, and eventually outcompete them. The authors proposed that a reduction in nodule oxygen permeability appears to be the primary mechanism of plant defence against uncooperative rhizobia, leading to a reduction in oxygen available for aerobic respiration, and thus a decrease in bacteroid growth rate relative to cooperative bacteroids. The layers of plant-

imposed control in the legume-rhizobia symbiosis have been recently reviewed by Porter et al. (2024).

The symbiosis of rhizobia with *Parasponia*, a tropical tree within *Cannabaceae*, appears to have arisen more recently than the legume-rhizobia symbiosis. The *Parasponia*-rhizobia symbiosis and its differences with legume-rhizobia symbiosis was reviewed by Behm et al. (2014). Nodules of *Parasponia* show fewer differences from lateral roots than those of legumes, with less specialisation of vascular bundles and thus less efficient transport of nutrients between partners. In contrast to legumes, in *Parasponia*, while intracellular, rhizobia remain in contact with the apoplast of nodule cells. The number of diazotrophs per nodule cell and the efficiency of nitrogen fixation per symbiotic bacterium is also lower than in legumes. Interestingly, *Parasponia* can host a broad range of rhizobial species (Op Den Camp et al., 2012). Exhibiting less signalling specificity and less developmental reprogramming than legumes, *Parasponia* may be an interesting and more primitive guide for future engineering efforts to extend rhizobial symbioses to non-legumes.

1.2.2.2 Extending nodulation in non-legumes

While true nodulation has not been achieved outside native host of rhizobia, there have been some successes in the artificial induction of nodule-like structures to host diazotrophs in the cereals. In future work, the genetics underlying spontaneous nodulation in *Lotus japonicus* mutants (Tirichine et al., 2007) may prove helpful in transferring nodulation to non-cereals; this would however require the transfer of the whole signalling network governing Nod-dependent developmental reprogramming.

Tchan et al. (1991) have reported the formation of tumour-like growths they termed *para*-nodules on the root system of wheat (*Triticum aestivum* L.) after induction with the plant hormone auxin 2,4-dichlor-phenoxy acetic acid (2,4-D) during germination. Both Tchan et al.

(1991) and Christiansen-Weniger (1992) have found that acetylene reduction in inoculated roots was significantly higher in plants treated with 2,4-D; acetylene reduction activity was also more robust to oxygen exposure in the treated plants even as bacterial density on the root surface remained the same (Christiansen-Weniger, 1992). Christiansen-Weniger & Vanderleyden (1994) have successfully replicated this association in maize (*Zea mays*). In this work, the authors have confirmed the intracellular colonisation of *para*-nodules by *A. brasilense*. Additionally, Tchan et al. (1991) suggest that carbon supplied by the plant partner was sufficient to sustain the colonising bacterium without carbon supplementation.

1.2.2.3 Nodulation-independent Symbioses

Nitrogen-fixing cyanobacteria can establish nitrogen-supplementing symbiotic associations with a wider range of plant species than rhizobia. Cyanobacteria of the genus *Nostoc* (including many *Anabaena* species) are especially amenable to symbiotic associations, forming associations with a wide range of plant species, as well as a diversity of fungi in the form of lichens (Rikkinen et al., 2002). *Nostoc spp.* can also form symbioses with various bryophytes, notably *Azolla spp.*, cycads, and the dicot angiosperm *Gunnera spp.* (Meeks & Elhai, 2002). Colonisation of plant tissues by symbiotic cyanobacteria is associated with developmental reprograming of the cyanobacterial filament, resulting in increased frequency of nitrogen-fixing cells (heterocysts) (Hill, 1975). Interestingly, the symbiosis between plants and their respective cyanobionts is not obligate, as axenic plants have successfully been cultivated in nitrogen-supplemented media, though cyanobiont-free plants have not been reported in nature (Hill, 1977).

The nitrogen-fixer *Gluconacetobacter diazotrophicus* (species originally named *Acetobacter diazotrophicus*) was found to facultatively associate with sugarcane without requiring nodulation (Reinhold-Hurek & Hurek, 1998). Rates of *G. diazotrophicus* colonisation of sugarcane inversely correlate with soil nitrogen content (Fuentes-Ramirez et

al., 1993). Cocking et al. (2006) have demonstrated extensive intracellular colonisation of the plant model organism *Arabidopsis thaliana* as well as the following crops by *G. diazotrophicus*: maize (*Zea mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), oilseed rape (*Brassica napus*), tomato (*Lycopersicon esculentum*), and white clover (*Trifolium repens*) under laboratory conditions. Expression of nitrogenase by the colonising bacterium was confirmed. In a field study by Ceballos-Aguirre et al. (2023), the application of *G. diazotrophicus* to carrot (*Daucus carota* L.) fields 30 days after sowing was associated with a 34% improvement in yield without additional nitrogen supplementation; the authors did not establish whether plant internal colonisation was achieved in this case.

The application of *Rhizobium leguminosarum* bv. *trifolii* – the rhizobium normally associated with clover (*T. repens*) – in combination with N fertiliser has been shown to increase grain yield of several varieties of rice (*Oryza sativa* L.) in a large-scale field trial by an average of 19.5% relative to N fertiliser alone; this trial did not investigate whether the rhizobium has become endophytic, or remained within the soil / root surface only (Yanni & Dazzo, 2010). Reddy et al. (1997) have demonstrated that rice root exudate does not activate the expression of nodulation genes in the rhizobia species *Bradyrhizobium japonicum* USDA110, *R. leguminosarum* bv. *trifolii* or *R. sp.* NGR234. In the same study, neither the developmental reprogramming of root hairs required for *para*-nodules or true nodulation could be brought on by purified Nod factors or inoculation with rhizobia, though increased formation of morphologically normal root hairs was reported. The authors suggest that the infection of rice by rhizobia is Nod-independent and does not proceed in a manner similar to the specific infection of legumes. Infecting rhizobia were limited to intercellular spaces and dead plant cells in the superficial levels of the root, and no membrane-enclosed endosymbiotic rhizobia in healthy cells could be found. Inoculation of rice with rhizobia has been reviewed by (Abdullahi et al., 2020).

Research into transferring nodulation to non-cereals has largely been abandoned since the 1990s with little explanation to the reasons behind this in literature. The refocusing of research groups on the direct transfer of nitrogenase to plants with the emergence of modern genetic and protein engineering techniques may be one culprit. At the same time, elucidating a method to engineer the complex organism-level developmental changes required for the induction of nodulation may prove to be an interesting question to future generations of researchers in plant synthetic and developmental biology that could not be answered using techniques available at the time.

1.2.3 Rhizosphere enrichment

1.2.3.1 Plant Growth-Promoting Rhizobacteria (PGPRs)

Up to 20% of carbon-containing compounds produced in photosynthesis, or photosynthate, is released into soil through the root system, providing a rich source of nutrients for the soil microbiome (Turner et al., 2013). The soil surrounding a plant's root system (rhizosphere) is rich with a diversity of microorganisms which benefit from this nutrient. Of these, bacteria beneficial to the plants are commonly referred to as plant growth-promoting rhizobacteria (PGPRs). PGPRs can improve crop outcomes through the production of protective antibiotics and enzymes, which help degrade environmental toxins, inducing systemic resistance, synthesis of volatile organic compounds and exopolysaccharides and iron-containing siderophores. They can play a role in heavy metal detoxication, improved salinity tolerance and insect control. PGPRs can also directly promote plant growth by increasing the supply of nutrients through nitrogen fixation, solubilisation of phosphate and potassium from inorganic sources, the production of plant hormones (phytohormones) (Singh et al., 2019). A diversity of the agricultural benefits application of PGPRs has been reviewed by Vessey (2003).

Free-living nitrogen fixers (FLNFs) are a class of PGPRs ubiquitous in soil ecosystems, where they consume carbon released by plant root exudates and enrich soil with secreted bioavailable nitrogen. FLNFs include bacteria of the genera *Azospirillum*, *Burkolderia*, *Gluconacetobacter*, *Pseudomonas*, *Azotobacter*, *Arthrobacter*, *Acinetobacter*, *Bacillus*, *Enterobacter*, *Erwinia*, *Flavobacterium*, *Klebsiella*, *Acetobacter, Anabaena, Nostoc* and others; many species can exist both as endophytes and FLNFs (Nagargade et al., 2018). Interestingly, they contribute to agriculture less than 10% of the nitrogen that symbiotic rhizobia do, even with comparable rates of nitrogen fixation and assimilation (Reed, 2011). Our understanding of the regulation of rhizosphere nitrogen fixation and the communication

between plants and soil microbiota is unfortunately limited, as research on rhizobia-like symbioses has overshadowed research into FLNFs (Smercina et al., 2019).

The enrichment of the plant rhizosphere, or the soil surrounding plant roots, with FLNFs has been successfully shown to improve the growth of crop plants including cereals. For example, nitrogen-fixing bacteria of the genus *Azospirillum*, notably *A. brasilense*, form associations with sugarcane with partial colonisation of plant interior, improving nitrogen accumulation and grain yield. These findings were later replicated in maize, wheat and pearl millet, suggesting that around 60% of the nitrogen demands of maize can be met through root inoculation with *Azospirillum sp.* (Boddey & Dobereiner, 1995). In a field study, maize seeds inoculated with *A. brasilense* and diazotroph *Azorhizobium caulidonans*, normally an endophyte of the legume sesbania (*Sesbania rostrata*), showed increased shoot dry weight and grain yield in the absence of additional nitrogen supplementation; this effect was only apparent in nitrogen-deplete, and not in nitrogen-rich soil (Woodard & Bly, 2000).

Research by Lupwayi et al. (2004) also suggests that one of the mechanisms by which the crop rotation, the traditional practice of growing multiple crops on the same land in a cyclical sequence, improves crop yields is by enriching the rhizosphere of non-leguminous plants within the sequence with nitrogen fixing bacteria, and not purely by supplementing nitrogen fixed during the legume growth phase.

Overall, the research into the practical applications of the interactions of free-living nitrogen fixers with plant crops is limited, although some industrial inoculants largely including wild-type bacteria as seed coatings are available commercially. There has been little work reported on improving the efficacy of nitrogen provision by free-living nitrogen fixers to plants, or on improving the amenability of plants to this bacterial nitrogen.

1.3 Designing Obligate Consortia with Carbon-Nitrogen Exchange Using New-to-Nature Exchange Molecules

Bioavailable nitrogen is an essential plant nutrient; however, biological nitrogen fixation is limited to a subset of prokaryotes. In this project, I present an approach to addressing the nitrogen crisis based on engineered photoautotroph-diazotroph symbioses. Specifically, this project focuses on the induction of a metabolic symbiosis using free-living nitrogen fixers.

We propose that the inefficiencies of nutrient provision by free-living nitrogen fixers to plants can be in part overcome by engineering an obligate symbiosis between the photosynthetic organism and its nitrogen-fixing partner. Up to 20% of plant photosynthate is released through the root system (Turner et al., 2013), providing the root ecosystems, the rhizosphere, with a natural source of nutrients making it rich with a diversity of bacteria, including diazotrophs. In spite of this, the impact of free-living rhizosphere diazotrophs on plant nutrient demands is negligible in comparison to the efficacy of symbiotic rhizobia of leguminous plants. The contribution of FLNFs to global agriculture is less than 10% that of symbiotic rhizobia, even with comparable rates of nitrogen fixation and assimilation (Reed, 2011).

I investigate the theoretical efficacy of a wide range of pathways for the exchange of nutrients between carbon-fixing and nitrogen-fixing members of a consortium acting as analogues for crops and rhizosphere diazotrophs, respectively. In each case, a nitrogen-carrying compound is synthesised in the diazotroph utilising nitrogen fixed from the atmosphere, transported into the photoautotroph, and broken down to be used as a nitrogen source. Analogously, a carbon-carrying compound is produced by the photoautotroph using photosynthesis-derived organic compounds and transported into the diazotroph to be used as a

source of energy and carbon backbones for biosynthetic processes. We propose that nutrient exchange pathways using new- or rare-to-nature nutrient-exchange compounds can be used to overcome some of the limitations associated with nutrient transfer between free-living nitrogen fixers and plants – namely the competition for resources imposed by other members of the ecosystem, as pathways breaking down more exotic compounds are less likely to have arisen and been selected for spontaneously. Nutrients synthesised by both partners are therefore more likely to remain within the consortium system, which is then favoured over organisms which cannot assimilate atmospheric carbon and nitrogen or the fixed carbon and nitrogen produced by the consortium.

First, I establish an *in vitro* consortium of the photoautotroph *Synechocystis sp.* PCC 6803, a model non-diazotrophic unicellular cyanobacterium, which acts as a stand-in for a plant system, with *Azotobacter vinelandii*, a fully aerobic nitrogen fixing bacterium ubiquitous in soil ecosystems. This consortium is subsequently also implemented as a genome-scale metabolic model *in silico*. This model captures the metabolic capabilities of both consortium partners and enables the evaluation of the effects of different nutrient exchange pathways on growth rates achievable by the consortium partners under nutrient-limited conditions using flux-balance analysis. Using the consortium model, I investigate the potential of compounds which are naturally synthesised by both partners to act as nutrient-carrying compounds. Additionally, a set of nutrient exchange pathways involving compounds central to nitrogen metabolism are tested – namely ammonia, glutamate, glutamine and α-ketoglutarate. This work is compared with the nutrient exchange system naturally found in the diazotrophic multicellular cyanobacterium *Anabaena sp.* PCC 7120.

Next, the nutrient-exchange compound design space is expanded using a high-throughput *in silico* investigation of compounds that do not normally exist within either consortium partner and may be underrepresented in natural ecosystems. I apply a

bioretrosynthesis algorithm (Koch et al., 2019) to evaluate the potential of *A. vinelandii* and *Synechocystis* 6803 to synthesise and break down all 1,048,228 carbon-containing and 381,616 nitrogen-containing compounds deposited in the MetaNetX database (Moretti et al., 2021). The total set of biosynthetic and breakdown pathways thus designed are subsequently implemented in a genome-scale metabolic model of the consortium to evaluate which pathways can carry flux and act as effective nitrogen or carbon-exchange compounds in the consortium in question.

We believe that a renewed focus on the optimisation of free-living diazotrophs as plant partners is necessary to complement modern efforts to address the nitrogen crisis alongside efforts to extend nitrogen fixation to plants and to engineer nodulation in non-legumes. This project is a step towards improving the symbioses of plants with rhizosphere nitrogen-fixing partners in which a high-throughput approach is used to address the inefficiencies of nutrient transfer between photoautotrophic and diazotrophic partners.

Chapter 2
Materials and Methods

2.1 Molecular Biology Methods

2.1.1 Amplification of DNA parts

E. coli DH5a genomic DNA (Genbank: CP033092.2) was imported into Benchling. Primers flanking parts to be amplified as bioparts were designed in Benchling. A BASIC prefix 5' TCTGGTGGGTCTCTGTCC 3' was added on an overhang of the forward primer, and a reverse complement of the BASIC suffix 5' CGATAGGTCTCCCGAGCC 3' was included on the reverse primer . All primers were ordered from IDT. Genomic DNA was purified from *A. vinelandii* DJ and *E. coli* K12 using the Monarch® Genomic DNA Purification Kit (New England Biolabs) following manufacturer's instructions. Bioparts were amplified using Q5 DNA polymerase per Table 2, Table 3 and Table 4.

Table 2: Reaction mixture for amplification of bioparts from genomic DNA

Component	**Volume**
5x Q5 Reaction buffer	10 µL
10mM dNTPs	1 µL
10µM forward primer	2.5 µL
10µM reverse primer	2.5 µL
Q5 high fidelity DNA polymerase	0.5 µL
5x Q5 high GC enhancer	10 µL
Template (gDNA)	1 µL
Nuclease-free water	22.5 µL

Table 3: Cycling protocol for the amplification of bioparts from genomic DNA for cloning

Number of cycles	Step	Temperature	Duration
	Initial denaturation	98°C	30s
25 cycles	Denaturation	98°C	5s
	Annealing	Table 4	10s
	Extension	72°C	40s
	Final extension	72°C	2min

Table 4: Annealing temperatures for the amplification of bioparts from genomic DNA for cloning

Primer pair	Annealing temperature
AV_GDH_downstream_F_BASIC_prefix AV_GDH_downstream_R_BASIC_suffix	57°C
AV_GDH_upstream_F_BASIC_prefix AV_GDH_upstream_R_BASIC_suffix	57°C
AV_GOGAT_upstream_F_BASIC_prefix AV_GOGAT_upstream_R_BASIC_suffix	57°C
R_Avinelnadii_scrT_BASIC_suffix F_Avinelnadii_scrT_BASIC_prefix	59°C
F_K12_yblJ_BASIC_prefix R_K12_yblJ_BASIC_suffix	58°C
R_scrY-1_BASIC_suffix F_AV_scrY_SDM_BsaI_removal	58°C
F_scrY-1_BASIC_prefix	58°C

R_AV_scrY_SDM_BsaI_removal	
F_AV_scrY_SDM_BsaI_removal R_AV_scrY_SDM_BsaI_removal	60°C
AV_GOGAT_downstream_R_BASIC_suffix F_AV_GOGAT_downstream_SDM_BsaI_removal	56°C
AV_GOGAT_downstream_F_BASIC_preffix R_AV_GOGAT_downstream_SDM_BsaI_removal	56°C
F_AV_GOGAT_downstream_SDM_BsaI_removal R_AV_GOGAT_downstream_SDM_BsaI_removal	58°C

2.1.2 Amplification of parts requiring the removal of an internal BsaI sites

Overlap-extension PCR (OE-PCR) was used for the removal of internal BsaI restriction enzyme sites where required. An upstream and downstream primer carrying BASIC prefix and suffix were designed like in 2.1.1. Two internal primers both covering the BsaI recognition site were designed to introduce a silent point mutation removing the restriction sites (Figure 7). A first amplification using *A. vinelandii* DJ gDNA as template was performed using primers F1 paired with R2, and F2 paired with R1, conditions equal as in Table 2 and Table 3, except only using 15 PCR cycles (for primer sequences, see Supplementary table 3). The PCR product was purified QIAquick PCR Purification Kit (QIAGEN) using 15 µL MiliQ water to elute, and used as template in a second round of PCR (Table 3, Table 5).

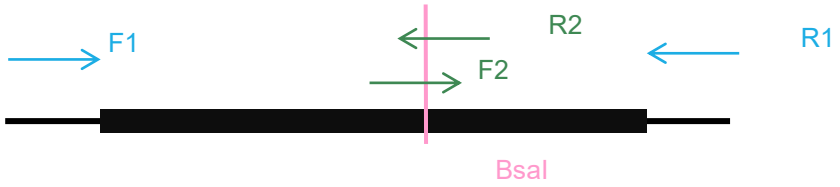


Figure 7: Schematic representation of the layout of primers in OE-PCR for the amplification of bioparts while removing internal BsaI restriction sites. Black rectangle designates genomic region of interest. Blue arrows are outside primers carrying BASIC prefix and suffix sequences. Internal primers are shown in green, overlapping the BsaI recognition site in pink.

Table 5: Reaction mixture for the 2nd amplification of bioparts requiring the removal of an internal BsaI site from genomic DNA

Component	**Volume**
5x Q5 Reaction buffer	10 µL
10mM dNTPs	1 µL
10µM forward primer	2.5 µL
10µM reverse primer	2.5 µL
Q5 high fidelity DNA polymerase	0.5 µL
5x Q5 high GC enhancer	10 µL
Template fragment 1	1 µL
Template fragment 2	1 µL
Nuclease-free water	21.5 µL

2.1.3 Gel electrophoresis

All amplified parts were validated using agarose gel electrophoresis in 1% agarose at 90-120 V for 20-60 min and imaged under UV light, unless otherwise specified.

2.1.4 DNA synthesis

BASIC assembly (Storch et al., 2015) compatible parts were ordered from IDT either as gBlocks or as cloned parts. Internal BsaI sites were manually removed from sequences in the Benchling web application (Benchling, 2021). All parts were flanked the upstream prefix 5'-TCTGGTGGGTCTCTGTCC-3' and downstream suffix 5'-GGCTCGGGAGACCTATCG-3'. With the exception of homology arms that necessitate high sequence identity, parts were codon optimised for the recipient organism using the IDT codon optimisation tool (https://eu.idtdna.com/CodonOpt).

2.1.5 DNA part subcloning using CloneJET

DNA parts synthesised using PCR or received as gBlocks were subcloned into the pJET plasmid for storage using the CloneJET PCR cloning kit (Thermo Scientific). Cloning was carried out using a modified protocol provided by fellow lab member Dr Jonathan Muller as follows: The reaction mixture (Table 6) was incubated for 30 min at room temperature, followed by transformation into calcium competent *E. coli* DH5α using heatshock (Section 2.2.1.3).

Table 6: Modified reaction mixture for blunt end cloning using the CloneJET kit

Component	**Volume**
2x reaction buffer	2 µL
pJET1.2 blunt	0.2 µL
DNA fragment (gBlock or PCR fragment)	3 µL
T4 DNA ligase	0.5 µL

2.1.6 DNA part subcloning using OE-PCR

Subcloning using overlap-extension PCR (OE-PCR) was used when colonies could not be obtained using the CloneJET kit (Section 2.1.5). The method for subcloning using OE-PCR was modified from (Bryksin & Matsumura, 2010). Chimeric primers were designed for subcloning of gBlocks using OE-PCR, in which a section of each primer anneals to the backbone (pJET 1.2), and a section to the insert to be subcloned (Figure 8). Overlaps with the insert were designed to overlap with the standard BASIC prefix and suffix sequences, allowing the same primer pair to be used for all gBlocks. For primer sequences, see Supplementary table 3.

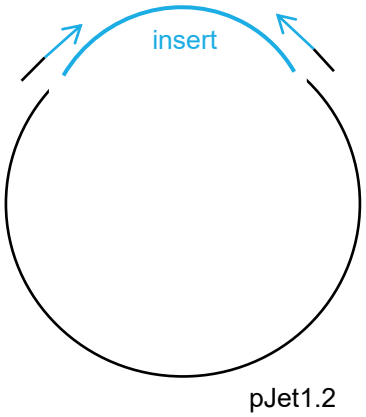


Figure 8: Schematic representation of primers used in OE-PCR-mediated subcloning. Primers are shown as arrows, where corresponding colours are used to show where on the backbone a section of the primer anneals.

The first round of PCR was carried out using synthesised primers and the gBlock as template. The stage 1 cycling protocol was designed to have 10 cycles at a lower annealing temperature, accounting for the annealing temperature of primers with overhangs, followed by

20 cycles at a higher annealing temperature, corresponding to the annealing temperature once all template contains the full overhang and the whole primer may anneal (Table 7 and Table 8).

Table 7: Reaction mixture for DNA part subcloning using OE-PCR, stage 1

Component	**Volume**
5x Q5 Reaction buffer	10 µL
10mM dNTPs	1 µL
10µM forward primer (F_OEPCR_BASIC_into_pJET)	1.25 µL
10µM reverse primer (R_OEPCR_BASIC_into_pJET)	1.25 µL
Q5 high fidelity DNA polymerase	0.25 µL
5x Q5 high GC enhancer	5 µL
Template (10ng/µL gBlock)	0.5 µL
Nuclease-free water	19.25 µL

Table 8: Cycling protocol for DNA part subcloning using OE-PCR, stage 1

Number of cycles	Step	Temperature	Duration
	Initial denaturation	98°C	30s
10 cycles	Denaturation	98°C	10s
	Annealing	60°C	30s
	Extension	72°C	90s
20 cycles	Denaturation	98°C	10s
	Annealing	65°C	30s
	Extension	72°C	90s
	Final extension	72°C	2min

DNA from stage 1 was next purified using QIAquick PCR Purification Kit (QIAGEN) following manufacturer instructions and using 15 µL MiliQ water to elute. This was then used as a megaprimer for the second stage of OE-PCR, where the destination plasmid serves as template (Table 9 and Table 10).

Table 9: Reaction mixture for DNA part subcloning using OE-PCR, stage 2

Component	Volume
5x Q5 Reaction buffer	5 µL
10mM dNTPs	0.5 µL
Megaprimer	13.75 µL
Q5 high fidelity DNA polymerase	0.25 µL
5x Q5 high GC enhancer	5 µL
Template (pJet1.2 from CloneJET kit)	0.25 µL

Table 10: Cycling protocol for DNA part subcloning using OE-PCR, stage 2

Number of cycles	Step	Temperature	Duration
	Initial denaturation	98°C	2min
30 cycles	Denaturation	98°C	30s
	Annealing	60°C	1min
	Extension	72°C	2min
	Final extension	72°C	10min

2.1.7 BASIC Assembly

Bioparts Assembly Standard for Idempotent Cloning (BASIC) (Storch et al., 2015) was chosen to assemble DNA parts. Bioparts were first digested with BsaI and ligated with the appropriate prefix and suffix linkers in a one-tube reaction (Table 11) in a thermocycler (2 h at 37°, 20 min at 20°, 20 min at 65°, optional 4° hold).

Table 11: BASIC Digestion/Ligation Reaction Setup

Component	Volume
Promega T4 buffer	3 µL
iP Linker	5 µL
iS Linker	5 µL
DNA part	To a final concentration of ~6.7 ng/µL, typically 1 µL
NEB BsaI-HF-v2	1 µL
Promega T4 DNA ligase	0.5 µL
Nuclease-free water	To a total volume of 30 µL

Each digestion/ligation reaction product was next individually purified to remove unligated bioparts using AmpliClean magnetic beads (Nimagen) following manufacturer's instructions. Purified DNA was assembled (Table 12) at 50° for 45 min. Finally, the whole assembly reaction was used to transform chemically competent *E. coli* DH5α using heatshock (Section 2.2.1.3).

Table 12: BASIC Assembly reaction

Component	**Volume**
DNA part (each)	2 µL
NEB Cutstmart buffer	1 µL
Nuclease-free water	To a total volume of 10 µL

2.1.8 Colony PCR and gel electrophoresis from *E. coli* DH5α

Colonies were picked from overnight plates using a sterile pipette tip and resuspended in 10 µL nuclease-free water by pipetting up and down.

DNA was amplified using DreamTaq PCR Master Mix (2x) or DreamTaq Green PCR Master Mix (2x) (Fisher Scientific) using primers annealing to the pJET plasmid backbone, pJET_F and pJET_R, or pJET1.2 Forward Sequencing Primer and pJET1.2 Reverse Sequencing Primer (Table 13 and Table 14). Amplified fragments were later analysed on 1% agarose gel, or 4% agarose gel for fragments < 500 bp. See Supplementary table 3 for primer sequences.

Table 13: Reaction mixture for colony PCR

Component	**Volume**
DreamTaq PCR Master Mix (2x)	12.5 µL
10mM forward primer (pJET_F)	1.25 µL
10mM reverse primer (pJET_R)	1.25 µL
Cells resuspended in nuclease-free water	10 µL

Table 14: Cycling protocol for colony PCR

Number of cycles	**Step**	**Temperature**	**Duration**
	Cell lysis and initial denaturation	95°C	5 min
25 cycles	Denaturation	95°C	1 min
	Annealing	52°C	1 min
	Extension	72°C	2 min
	Final extension	72°C	10 min

2.1.9 Sequencing from *E. coli* DH5α

Colonies were picked from overnight plates using a sterile pipette tip to inoculate 5mL liquid LB culture in a 50mL Falcon tube. Plasmid DNA was isolated from overnight cultures using QIAprep Spin Miniprep Kit (QIAGEN) following manufacturer instructions. DNA concentration was measured using Tecan Infinite 200 PRO NanoQuant Plate in Tecan Infinite M200 Microplate Reader (Tecan). Samples were submitted for Sanger sequencing by Source Biosciences using primers annealing to the pJET plasmid backbone, pJET_F and pJET_R

(Supplementary table 3). Sequencing results were analysed in Benchling (Benchling, 2021). Positive colonies were then preserved as glycerol cryostocks.

2.2 Cell Culture and Transformation

2.2.1 *E. coli* DH5α

Escherichia coli DH5α was obtained from Dr Daniel Zabala Alvarez in Prof Patrik Jones's group. Luria-Bertrani (LB) medium (Bertrani, 1951) with appropriate antibiotic (Supplementary table 10) was used to culture *E. coli* DH5α. Unless otherwise specified, liquid cultures were grown at 37°C and 150 rotations per minute (rpm). Overnight cultures for cryopreservation (Section 2.2.1.1) or minipreps were inoculated using a sterile tip to pick colonies and grown in 5 mL media in 50 mL sterile falcons, without antibiotic. Solid cultures on LB agarose plates were grown inverted at 37°C. When required, plates sealed with parafilm were stored inverted at 4°C.

2.2.1.1 Preparation of and Retrieval from Glycerol Stocks

Stocks for cryopreservation were prepared from overnight cultures. 500 µL of culture was combined with 500 µL autoclaved 50% glycerol in MiliQ water in a cryotube, catalogued, and stored labelled at -80°C.

To streak out from cryopreservation, the contents of the cryotube were scraped with a sterile inoculation loop in a sterile flow hood environment and transferred to a BG11 plate with appropriate antibiotic (Supplementary table 10). The cryotube was kept on ice to avoid freeze-thaw cycles and returned to -80°C as soon as possible. The plate was then incubated normally.

2.2.1.2 Preparation of Chemically Competent Cells

To prepare chemically competent *E. coli* DH5α, a single colony from an overnight LB plate was picked and inoculated in a 10 mL starter culture in liquid LB and grown overnight at 160 rpm, 37°C. The next day, 1 L LB media was inoculated with 10 mL starter culture and grown at 160 rpm, 37°C and for approximately 3 h to OD_{600} 0.35-0.4. Cells were chilled on ice

for 30 min and occasionally manually swirled. 1 L culture was next split into 50 mL aliquots in pre-chilled 50 mL Falcon tubes and harvested by centrifugation at 3000 g, 15 min, 4°C. The pellet was resuspended in 20 mL ice-cold 100 mM $MgCl_2$. Cells were then harvested by centrifugation at 2000 g, 15 min, 4°C. The pellet was resuspended in 40 mL ice-cold 100 mM $CaCl_2$, and chilled on ice for 20 min. Cells were harvested by centrifugation at 2000 g, 15 min, 4°C. The pellet was resuspended in 10 mL ice-cold 85 mM $CaCl_2$ with 15% glycerol. Cells were harvested by final centrifugation at 1000 g, 15 min, 4°C and the pellet was finally resuspended in 0.5 mL ice-cold 85 mM $CaCl_2$ with 15% glycerol. 100 µL aliquots were stored in cryotubes at -80°C. Sterile technique was used throughout.

2.2.1.3 Heatshock Transformation

100 µL of chemically competent *E. coli* DH5α were thawed on ice. Cells were incubated with 10 µL cloning reaction for 30 min on ice. The mixture was next heat-shocked at 42°C for 60 s and cooled on ice for 2 min. 500 µL of SOC pre-heated to 37°C was added, and the whole mixture was incubated at 37°C in a shaking incubator for 15 min when using ampicillin or carbenicillin, and 1 h for other antibiotics. Cells were then plated on LB plates with appropriate antibiotic (Supplementary table 10) and grown overnight at 37°C. Sterile technique was used throughout.

2.2.2 *Synechocystis sp.* PCC 6803

Synechocystis sp. PCC 6803 was obtained from Dr Marine Valton in Prof Patrik Jones's group. Unless otherwise specified, cells were grown in BG11 medium (Supplementary Methods 9.2.2.2) in 250 mL culture flasks at 160 rpm at 30°C, 1% CO_2 and 60 µE, or on BG11 medium agar plates at 30°C, 1% CO_2 and 60 µE illumination in a PSI Algaetron AG 230 growth chamber.

2.2.2.1 Preparation of and Retrieval from Glycerol Stocks

Stocks for cryopreservation were prepared from overnight cultures. 500 µL of culture was combined with 500 µL autoclaved 50% glycerol in MiliQ water in a cryotube, catalogued, and stored labelled at -80°C.

To streak out from cryopreservation, the contents of the cryotube were scraped with a sterile inoculation loop in a sterile flow hood environment and transferred to a BG11 plate with appropriate antibiotic. The cryotube was kept on ice to avoid freeze-thaw cycles and returned to -80°C as soon as possible. The plate was incubated at 30°C, 1% CO_2 covered with tissue for two days to mitigate photobleaching, and normally after this. Sterile technique was used throughout.

2.2.3 A. vinelandii

A. vinelandii protocols and media formulations were largely adapted from Dos Santos (2011). *A. vinelandii* DJ was obtained by Dr Marine Valton in Dr Patrik Jones's group from Barney et al. (2015). Burke's BN medium (Supplementary Methods 9.2.2.1) (Dos Santos, 2011) was used to culture *A. vinelandii*. Unless otherwise specified, liquid cultures were grown at 30°C, 150 rpm, in baffled bottom flasks, or in 6 well plates 160 rpm at 30°C, 1% CO_2 and 60 µE in a PSI Algaetron AG 230 growth chamber for consistency with *Synechocystis* 6803 and coculture experiments. Solid cultures on BN media plates were grown inverted at 30°C. When required, plates sealed with parafilm were stored inverted at 4°C.

2.2.3.1 Preparation of Competent Cells

Two consecutive passages of cells on BN plates lacking molybdenum (BN-Mo), grown for 3 days at 30°C were prepared. One passage of cells was next grown on BN plates lacking molybdenum and iron (BN-Mo-Fe) for 3 days at 30°C. If colonies are noticeably green at this point, a liquid culture can be inoculated; otherwise, an additional passage on BN-Mo-Fe is

required. Once green, a small loopful of cells is used to inoculate a 50mL BN-Mo-Fe liquid culture in 125 mL baffled flasks, and incubated at 30°C, 170 rpm for 18-20 h. Cells can only be used for transformations once bright green.

2.2.3.2 Natural Transformation

10 µL assembly mixture (from section 2.1.7) was mixed with 200 µL fresh competent cells and 200 µL transformation buffer (20mM MOPS buffer, pH 7.2, with 20 mM $MgCl_2$, filter sterilised), incubated for 20 min at room temperature, followed by spreading on solid selective BN media (Dos Santos, 2011).

2.2.4 Coculture of *A. vinelandii* and *Synechocystis* 6803

A sterile cell culture insert for 6-well plates with 0.4 µm pore, translucent high-density PET membrane (Falcon 353493) was placed inside a well of a 6-well plate. 3 mL of media was slowly pipetted into the insert, waiting until medium distributes both outside and inside the insert. 1 mL of *A. vinelandii* culture was next pipetted under the insert at to final OD680 of 0.01 and 1 mL of *Synechocystis* 6803 culture to final OD730 of 0.1 was pipetted into the insert. Where only one inoculant was used, the partner was replaced with 1 mL of media to a final volume of 5 mL per well. 6-well plates were carefully sealed with microporous tape (3M) allowing for gas exchange. When fewer than 6 wells were used in a plate, the extra wells were filled with media to ensure even evaporation. 6-well plates were cultured at 160 rpm at 30°C, 1% CO_2 and 60 µE in a PSI Algaetron AG 230 growth chamber.

2.2.5 Growth Assays

Unless stated otherwise, all growth assays were conducted in three wells as three biological triplicates, with three independent repeats in separate 6-well plates as three technical triplicates.

2.2.5.1 Calibration Curves

A set of 6 mL culture samples of each *A. vinelandii* and *Synechocystis* 6803 was diluted with MiliQ water to arbitrarily determined different concentrations. While this work (Figure 10) suggests there is no real difference in the two calibration curves for either species examined, convention was still followed for consistency and reproducibility. Thus, when recalculating from OD to gDW/L, the optical density at 680 nm (OD680) was used for *A. vinelandi* and at 730 nm (OD730) for *Synechocystis* 6803. All OD measurements were carried out on the same plate reader to avoid inconsistencies between machines. From each, OD680 and OD730 of a 200 µL aliquot was measured in a 96-well clear flat bottom plate (Falcon) using a Tecan Infinite M200 Pro plate reader.

A NC45 Whatman nitrocellulose membrane filter (Sigma-Aldrich) was dried for 15 min in a 60°C kiln and weighed using an analytical balance. The filter was secured on a vacuum filtration apparatus and loaded with 5 mL of culture sample and washed with MiliQ water to remove all liquid. After filtration, the filter was dried in a kiln for at least 3 h for complete drying and weighed again after.

Unloaded filter weight was subtracted from the loaded filter weight to determine cell dry weight in a 5 mL sample, which was subsequently recalculated for grams of dry weight per litre (gDW/L). A line of best fit (least squares) was fit for gDW/L against OD680 and OD730 in Microsoft Excel. The line equation was used to recalculate OD to gDW/L.

2.2.5.2 Preculture Washing and Medium Change

The OD of preculture was measured. 50 mL of preculture was first harvested by centrifugation at 3600 rcf, 10 min, room temperature. Under sterile conditions, the supernatant was gently poured out, replaced with sterile MiliQ water, and centrifuged again. The supernatant was poured out and the pellet was resuspended to appropriate OD with new fresh media.

2.2.5.3 Sample Collection and OD Measurement

Three 200 µL aliquots were pipetted from each well of a 6-well plate. When using well inserts, three 200 µL aliquot was taken from both inside the insert and outside of it for an initial OD reading. OD was measured in a 96-well clear flat bottom plate (Falcon) using a TECAN plate reader. OD readings were recalculated into grams of dry weight per litre and compared using unpaired Student's t-test.

Fresh media to a volume equal to volume taken out plus 137 µL to account for daily evaporation (Supplementary Methods 9.2.1) was finally added to culture for continued cultivation.

2.2.6 Quantification of Glutamate and Glutamine

2.2.6.1 Derivatisation with O-Phthaldehyde (OPA)

Derivatisation using o-phthaldehyde (OPA) with sodium sulfite was adapted from Monge-Acuña & Fornaguera-Trías (2009); (Zandy et al., 2017). Fresh derivatisation reagent (0.0055g OPA, 97% minimum purity dissolved in 125 µL ethanol absolute; 205 µL aqueous 1 M solidum sulphite, 98% minimum purity; 3.93 mL 0.1 M aqueous sodium tetraborate adjusted to pH 10.4 with 5 M sodium hydroxide) was prepared immediately before the HPLC run and kept at 12°C in the autosampler chamber. Online derivatisation was performed using an autosampler program when injecting (Table 15). An open vial of MiliQ water was used as injector wash.

Table 15: Autosampler program for online OPA derivatisation
Draw 2 µL derivatisation mix
Draw 18 µL sample
Mix 20 µL in air at default speed 10x
Wait 10 min
Inject 20 µL
Wash

2.2.6.2 HPLC Protocol

Glutamate and glutamine were separated and detected using an Agilent 1200 series HPLC with a Zorbax Eclipse XDB-C18 column (Agilent) (Table 17, Table 16). The method used was adapted from (Henderson & Brooks, 2010; Henderson et al., 2000; Shafaei et al., 2017; Wang et al., 2010). All solvents used were HPLC grade. Agilent Chemstation software was used for the viewing of peaks, graphing of spectra, and peak integration.

Table 16: HPLC setup details

Column	Zorbax Eclipse XDB-C18 (150 x 4.6 mm, 5 µm, Agilent 993967-902)
Mobile phase A	40 mM Na_2HPO_4 pH 7.8 [5.5 g $NaH2PO_4$ monohydrate in 1 L water, adjusted to pH 7.8 with 10 N NaOH) (Make 10x stock of mobile phase A to keep between runs, pH adjust for each new 1x solution)
Mobile phase B	Acetonitrile : Methanol : water (45:45:10, v/v/v)
Flow rate	1 mL/min
DAD Detection	Signal A: 338 nm, 10 nm bandwidth Signal B: 210 nm, 16 nm bandwidth
Peak width	> 0.03 min for the Traditional High Resolution 5-µm column methods
Injection volume	20 µL
Column oven	40 °C

Table 17: HPLC flow protocol

Time (min)	**% mobile phase B**
0	10
10	18
15	24
21	41
22	42.2
25	58
27	59
30	60

2.3 Computational Methods

Computational experiments were performed either on a home workstation running Windows 10 Pro, Intel(R) Core(TM) i7-4770 CPU @ 3.40 GHz processor with 24 GB RAM or using high-performance computing resources provided by the Imperial College Research Computing Service (http://doi.org/10.14469/hpc/2232) accessed via Windows Subsystem for Linux 2. This project uses Python 3.6 and 3.11 in Anaconda or Miniconda (Anaconda, 2024) and Matlab R2019a. The following Python packages are used throughout:

Table 18: Python packages used in this study

Python package	**Reference**
Equilibrator API	(Beber et al., 2022; Flamholz et al., 2012; Noor et al., 2012; Noor et al., 2014; Noor et al., 2013).
JSON	(Pezoa et al., 2016)
NumPy	(Harris et al., 2020)
Open Babel	(O'Boyle et al., 2011)
pandas	(McKinney, 2010)
PubChemPy	(Kim et al., 2023)
rdkit	(*RDKit: Open-source cheminformatics*; Laudrum & others, 2024)
RetroPathRL	(Koch et al., 2019)
Seaborn	(Waskom, 2021)
UMAP	(McInnes et al., 2018)
Matplotlib	(Hunter, 2007)

UMAP	(Healy & Mcinnes, 2024; McInnes et al., 2018)
hdbscan	(Mcinnes et al., 2017)

2.3.1 Retrieval of Candidate Compounds from Database

Candidate nutrient exchange compounds were retrieved from the MetaNetX database (Ganter et al., 2013; Moretti et al., 2016a; Moretti et al., 2016b; Moretti et al., 2021) files chem_xref.tsv and chem_prop.tsv at https://www.metanetx.org/mnxdoc/mnxref.html.

2.3.2 Conversion of Compound Representations

Compounds retrieved from MetaNetX database (Ganter et al., 2013; Moretti et al., 2016a; Moretti et al., 2016b; Moretti et al., 2021) in Section 2.3.1 were subset to exclude compounds labelled "fantasy" and include only compounds where column "#source" contained "biggM" – i.e. entries representing compounds linked to their BiGG identifiers, the format used for genome-scale metabolic models. Metabolites in BiGG model format were retrieved from http://bigg.ucsd.edu/data_access, file bigg_models_metabolites.txt (King et al., 2016).

The sets were then merged on identical BiGG identifiers, creating a table with MNXref (MetaNetX unique identifier), formula, InChI, InChiKey, SMILES, BiGG and human-readable names. The resultant table was used when converting between BiGG and InChI identifiers. See Supplementary file S3 for the full table. The rdkit python package was used to perform all other compound representation conversions where necessary.

2.3.3 Stoichiometric Modelling

The modification and analysis of stoichiometric models was carried out using Cobra Toolbox 3.0 in MATLAB R2019a or Cobrapy with Python 3.9.16 in Spyder 5.4.2 IDE with gurobi solver.

Table 19: Cobra Toolbox functions used

Function	**Usecase**
`changeObjective()`	Change objevtive function
`addReaction()`	Add reaction to model
`changeRxnBounds()`	Change bounds of reaction
`readCbModel()`	Load model from disk
`writeCbModel()`	Write model to disk
`optimizeCbModel()`	Find optimum solution
`verifyModel()`	Check model consistency
`setMediumConstraints()`	Specify medium constraints

2.3.4 Computational Bioretrosynthesis of Candidate Compounds

RetroPathRL was used to predict putative biosynthetic pathways for a full set of 381,616 nitrogen-containing and 1,048,228 carbon-containing compounds derived from the MetaNetX database (Bernard et al., 2014; Ganter et al., 2013; Moretti et al., 2016a; Moretti et al., 2021). All input compounds were formatted as InChI. The RetroPathRL configuration used is outlined in Table 20.

Configuration for individual RetroPathRL jobs was set up using a simple python script, creating a string representation of command line input (Code Block 1). This was subsequently

executed in batch mode on a high-performance computing cluster (http://doi.org/10.14469/hpc/2232).

Table 20: Parameter Choice for RetroPathRL. Adapted from RetroPathRL GitHub documentation (Koch, 2019).

Parameter	**Parameter Description**	**Value**
itermax	Maximum number of iterations of tree search	1000
expansion_width	Maximum number of children per node	10
time_budget	Maximum allotted time	7200
max_depth	Maximum depth of tree / maximum pathway steps	7
UCT_policy		Biochemical_UCT_1
UCTK	the constant defining exploration / exploitation in UCT formula	20
bias_k	weighing of progressive bias (if used)	0
k_rave	weighing of RAVE (if used)	0
Rollout_policy		Rollout_policy_random_uniform_on_biochemical_multiplication_score
max_rollout		3
chemical_scoring		SubandprodChemicalScorer
virtual_visits		0
progressive_bias_strategy		0

diameter	10 12 14 16
biological_score_cut_off	0.1
substrate_only_score_cut_off	0.5
chemical_score_cut_off	0.5
minimal_visit_counts	1

```
script = """python ../Tree.py
--log_file tree.log
--itermax 1000
--expansion_width 10
--time_budget 7200
--max_depth 7
--UCT_policy Biochemical_UCT_1
--UCTK 20
--bias_k 0
--k_rave 0
--Rollout_policy
Rollout_policy_random_uniform_on_biochemical_multiplication_s
core
--max_rollout 3
--chemical_scoring SubandprodChemicalScorer
--virtual_visits 0
--progressive_bias_strategy 0
--diameter 10 12 14 16
--c_name """ + c_name + """
--c_inchi """ + c_inchi + """
--folder_to_save """ + folder + """
--biological_score_cut_off 0.1
--substrate_only_score_cut_off 0.5
--chemical_score_cut_off 0.5
--minimal_visit_counts 1
--organism_name none
--complementary_sink """ + sink
```

Code Block 1: Python string representation of command line input for individual RetroPathRL runs. Variables are highlighted in cyan. Variable “c_name” stores compound name in human-readable format. “c_inchi” is the InChI identifier of the compound. “folder” is the output folder. “sink” is a list of compounds (as InChI) present in the chassis. All other parameters are constant across runs.

2.3.5 Reaction Reformatting

Output from RetroPathRL was yielded in the form of JSON files describing reaction participants and reactions within a pathway. A python script was used to iterate through all outputs, converting data to a comma-separated values (*.csv) format to optimise storage. Next,

for each pathway, the participating compounds and reactions were extracted, and reactions were individually stoichiometrically balanced (Section 2.3.6).

2.3.5.1 Cobra

To convert from RetroPathRL representations into Cobra-compatible representations, a conversion table matching InChI and BiGG identifiers created from MetaNetX and BiGG database entries (Supplementary file S3) was used. For some InChI identifiers, a metabolite with differing protonation state had to be used, as a perfect match was unavailable. For compounds for which a conversion was not possible as no BiGG identifier was found to match any protonation form of the compound were represented using their InChIKey. InChIKey was chosen over InChI in this application due to its shorter length to improve memory usage and human legibility of resultant scripts.

2.3.6 Stoichiometric Balancing of Chemical Equations

Reactions derived from RetroPathRL output required stoichiometric balancing and the addition of cofactors.

The following steps were followed for every reaction:

1) If an Enzyme Commission (EC) number was associated with the reaction, Brenda database (Chang et al., 2021) entry on this EC number was extracted using the brendapy python package, saving information on cofactors known to participate in reactions catalysed with by enzymes of this EC number. Reaction equations were also saved. These steps could only be carried out if the EC number was complete, i.e. a set of four numbers separated by periods. If a reaction was associated with multiple EC numbers, all entries were collected and combined.

2) Atoms in all compounds within a reaction were counted with the help of the rdkit python package. This required conversion from InChI representation to an rdkit mol object. Care was taken to explicitly declare hydrogens.
3) Atom counts on the left-hand side (reactants) and right-hand side (products) were compared. If there was a discrepancy, any elements missing from one side of the reaction were added in the following formats:
    a. In the case of oxygen atoms, a water molecule was added instead.
    b. In the case of hydrogen atoms, if an additional water molecule had not already solved the elemental imbalance, one of $NADH^+$, $FADH_2$ or $NADPH^+$ (in order of priority) was added depending on which of these cofactors was present in the Brenda entry. If more than one cofactor was present, only one cofactor was added in order of priority as outlined earlier. The oxidised form of the appropriate cofactor was added on the other side of the reaction. If no cofactor was declared in the Brenda entry, an H^+ ion was added instead.
    c. In the case of phosphate, ATP hydrolysis to ADP was included in the reaction; GTP to GDP was used instead if this was indicated by cofactors included in Brenda entry.
    d. For all other unbalanced elements, their elemental form was included.
4) A matrix was constructed to represent a list of all elements within the reaction and their respective counts in each participating compound, including negative numbers for reactants and positive numbers in the case of products. This matrix was solved using numpy.linalg.lstsq() to find integer coefficients that balance the equation.
5) If cofactors that balance the equation were found, the equation was returned. If cofactors that balanced the equation were not found, first, if coenzyme A was found

in the list of coefficients or the alternative reactions within the Brenda entry, it was added; coenzyme A was not added if it was not specified in the database. Then, the script was allowed to recur using the modified list of reactants and products.

6) If a balanced equation was returned, the balanced equation was used and labelled as balanced going forward.
7) If cofactors that balance the equation could not be found in the second run of the balancing function, the original reaction equation as specified in RetroPathRL output was used instead, labelling the reaction as unbalanced.

2.3.7 Compound Clustering

InChI was converted to Morgan Fingerprints using the RDKit Python package (Laudrum & others, 2024; Rogers & Hahn, 2010). Uniform Manifold Approximation and Projection (UMAP) (Healy & Mcinnes, 2024; McInnes et al., 2018) and the associated python package (https://github.com/lmcinnes/umap) with hdbscan (Mcinnes et al., 2017) was used for clustering and dimensionality reduction.

```
import hdbscan
from rdkit import DataStructs, Chem
from rdkit.Chem import AllChem
import umap

fps = len(compounds)*[None]
for i, compound in enumerate(compounds):
        mol = Chem.MolFromInchi(compound)
        fp = AllChem.GetMorganFingerprintAsBitVect(mol, 2)
        fps[i] = fp
embeddings = umap.UMAP(random_state=60,
             n_components = 2,
             n_neighbors= 15,
             min_dist = 0.20,
             metric="jaccard", verbose=True
             ).fit_transform([fp    for    fp    in    fps],
             force_all_finite = False)
labels = hdbscan.HDBSCAN(min_samples=5,
         min_cluster_size=5,
         ).fit_predict(embeddings)
plt.scatter(embeddings[:,0],
            embeddings[:,1],
            c = labels)
```

Code Block 2: Simplified Python 3 implementation of the determination of Morgan fingerprints followed by clustering, colouring by cluster and plotting.

Chapter 3
Establishing Cocultures *in vivo* and *in silico*

3.1 Introduction

3.1.1 Rationale

The design of novel metabolic pathways must be considered within the context of the organism of choice. In this chapter, I set up an *in vitro* and *in silico* consortium of the photoautotrophic cyanobacterium *Synechocystis sp.* PCC 6803, acting as a simplified plant stand-in, and the diazotroph *Azotobacter vinelandii*. A coculture system supporting the growth and separate monitoring of both organisms is presented. In parallel, I developed a two-cell genome-scale metabolic model to enable *in silico* modelling of the metabolic capabilities of the consortium based on separate published models; this is then used to guide the design of nutrient exchange pathways in later chapters.

3.1.2 Aims

- Select appropriate diazotrophic and photoautotrophic organisms
- Develop culture media and conditions capable of supporting the growth of both organisms in the same vessel
- Modify published genome-scale metabolic models (GEMs) for both partners to reflect a consortium, containing a full set of metabolites, reactions, and compartments of both organisms
- Apply biologically relevant constraints to the consortium model
- Validate model predictions

3.1.3 Choice of Chasses

A model photosynthetic partner has been chosen as stand-in for a plant partner to allow for faster pathway establishment and evaluation. *Synechocystis sp.* PCC 6803 (further as *Synechocystis* 6803) is a photosynthetic unicellular cyanobacterium incapable of nitrogen fixation. It is a commonly used model organism in photosynthesis and cyanobacterial research, including in our research group.

The choice of a nitrogen-fixing partner is complicated by the oxygen sensitivity of nitrogenase. In this work, I will be working with the obligate aerobe *Azotobacter vinelandii* DJ whose high rate of respiration serves as an effective oxygen scavenging mechanism to protect the nitrogenase enzyme in an aerobic atmosphere (Poole & Hill, 1997). It is therefore easier to handle in the laboratory than most other diazotrophs and can become a viable coculture partner for the oxygen-producing cyanobacterium. This bacterium is also commonly found in the rhizosphere and is not a plant or human pathogen. *A. vinelandii* has previously been demonstrated to function as a nitrogen donor in coculture with photosynthetic organisms (Lorincz et al., 2010; Barney et al., 2015). The non-gummy DJ strain was chosen due to laboratory practicality (Dos Santos, 2011).

Klebsiella oxytoca, another rare oxygen-tolerant diazotroph, was briefly investigated as an alternative nitrogen-fixing partner. *K. oxytoca* is a common model of nitrogen fixation with a well-studied nitrogen fixation gene cluster, commonly found in association with wheat roots. It can easily be cultured in common laboratory conditions and is compatible with many genetic tools originally developed for *E. coli*, increasing its attractiveness in biotechnology (Oldroyd & Dixon, 2014; Geddes *et al.*, 2015; Jeong *et al.*, 1998; Temme, Zhao & Voigt, 2012; Cakmakci, Evans & Seidler, 1981; Burén, López-Torrejón & Rubio, 2018). While *K. oxytoca* can be cultured under aerobic conditions, its oxygen tolerance is dependent on the formation

of biofilms impermeable to oxygen, and nitrogen-fixation only occurs in cells thus protected. This makes it impractical for liquid coculture in aerobic conditions with the oxygen-producing *Synechocystis* 6803.

Both *A. vinelandii* and *Synechocystis* 6803 are genetically tractable, with sequenced and annotated genomes (Table 21) and established genetic modification techniques. A genome-scale metabolic model is also available for both species (Table 21), which will enable effective metabolic modelling as part of this work (Chapter 7).

Table 21: Genome and metabolic model accession numbers for strains used in this study

Species	**Annotated genome**	**Metabolic Model**
Synechocystis sp. PCC 6803	BA000022.2 (Kaneko et al., 1996; Kaneko et al., 1995)	iSynCJ816 (Joshi et al., 2017)
Azotobacter vinelandii DJ	CP001157.1 (Setubal et al., 2009)	iAA1300 (Alleman et al., 2021)

3.2 Establishing Cocultures *in vitro*

3.2.1 Choosing a Coculture Medium

Establishment of viable cocultures requires the selection of a culture medium suitable for both organisms. I have compared typical culture media used for both *A. vinelandii* and *Synechocystis* 6803 to design a medium suitable for their coculture. Burk's B medium (Supplementary Methods 9.2.2.1) is commonly used to grow *A. vinelandii*. This can also be supplemented with ammonium acetate (BN medium) or urea (BU) as an additional source of nitrogen in non-diazotrophic conditions (dos Santos, 2011). *Synechocystis* 6803 is commonly cultured in BG11 medium (Supplementary Methods 9.2.2.2) (Rippka, 1972); fixed nitrogen (sodium nitrate) may be removed to create BG11-0 (Table 22). A simple BNxBG11 combination medium was created by supplementing BN with BG11 trace minerals solution to satisfy the nutrient needs of both species (Table 23).

Axenic growth of both species in the medium formulations under investigation (Table 22) was evaluated in liquid culture. Both species were able to grow in both BN and BNxBG11 combination medium (Figure 9). The growth of *Synechocystis* 6803 in both BN and in the combination medium outperformed its growth in BG11. This is likely due to the presence of sucrose in medium as a more readily available source of energy than sugars from photosynthesis the organism would depend on in carbon-free BG11. *A. vinelandii* did not grow in BG11 liquid culture, as this lacks fixed carbon. Interestingly, the growth of *A. vinelandii* in the combination medium outperformed its growth in BN.

Table 22: Comparison of carbon and nitrogen sources in media for *A. vinelandii* and *Synechocystis* 6803

Medium	**Nitrogen source**	**Carbon source**
B medium	no N source	60 mM sucrose
BN medium	20 mM ammonium acetate	60 mM sucrose
BU medium	10 mM urea	60 mM sucrose
BG11 medium	1.76 M sodium nitrate	No C source
BG11-0 medium	No N source	No C source
BNxBG11 combination medium	20 mM ammonium acetate	60 mM sucrose

Table 23: Recipe for combination medium for the coculture of *A. vinelandii* and *Synechocystis* 6803. The addition of Bacto Agar is necessary to prepare solid culture plates.

Component	**Volume to make 500mL media**
BN phosphate buffer	450 mL
BN 10x salts solution	49 mL
BG11 trace minerals	1 mL
(Optional: Bacto Agar)	(7.5 g)

A) Synechocystis sp. PCC 6803 B) Azotobacter vinelandii DJ

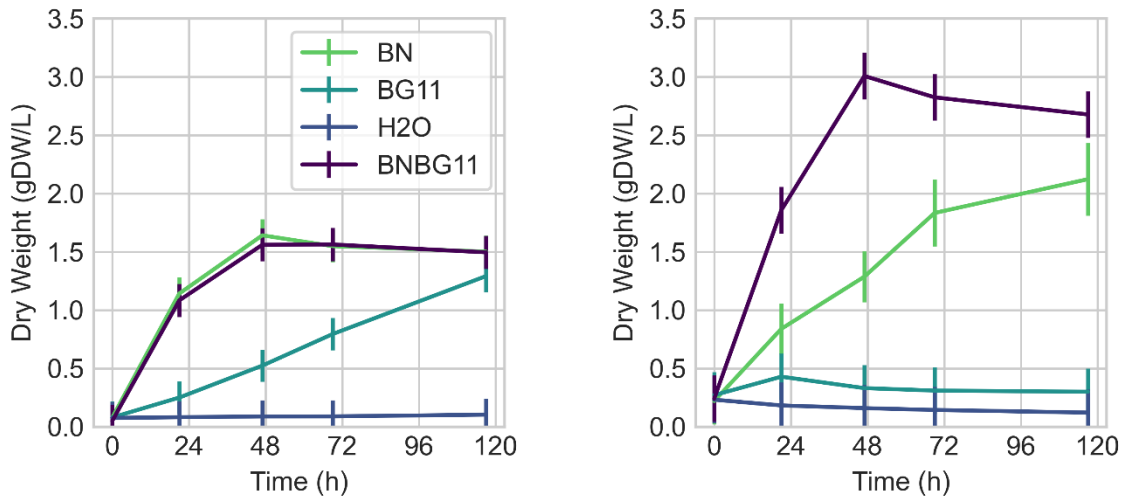


Figure 9: Growth curves of A) Synechocystis sp. PCC 6803 and B) Azotobacter vinelandii DJ in BN, BG11, combined media, and water as negative control. Error bars represent the combined standard error of the mean. (N = 3, n = 9)

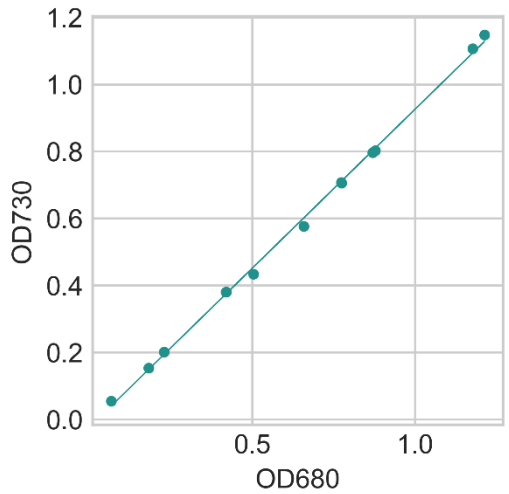
3.2.2 Developing a Method to Discriminate between Growth Rates of Coculture Partners

To monitor the composition of consortium culture over time, I needed to develop a method to track the growth rates of both species independently. Cell density is typically determined by proxy of culture optical density (OD) at 680 nm (OD680) for *A. vinelandii* and at 730 nm (OD730) for *Synechocystis* 6803. The higher wavelength used for photosynthetic organisms decreases overlap with chlorophyll absorbance. I have investigated whether the two wavelengths can be used to distinguish the species. The optical densities of cultures at 680 nm and 730 nm were compared (Figure 10). Interestingly, the very strong correlation between OD680 and OD730 ($R^2 \approx 1.00$ for both species) would suggest that the convention of using separate wavelengths for photosynthetic and non-photosynthetic species is unnecessary, although not detrimental. The high correlation between both measurements is likely due to the

optical density measurement corresponding to the scattering of light in solution, rather than true absorbance; scattering is therefore less likely to differ at relatively similar wavelengths. Future work could investigate the suitability of shorter wavelengths for *A. vinelandii* and *Synechocystis* 6803, as these are associated with lower scattering (Myers et al., 2013).

At the same time, a dry weight to optical density calibration curve was constructed for both species at both wavelengths (Figure 11). Both wavelengths were found to show a similar correlation to dry weight in both species, and therefore OD680:OD730 ratio cannot be used to evaluate the ratio of species in coculture. An approach utilising a physical barrier enabling the passage of media and, importantly, exchange metabolites, but not cells was investigated next.

A) *Azotobacter vinelandii* DJ



B) *Synechocystis sp*. PCC 6803

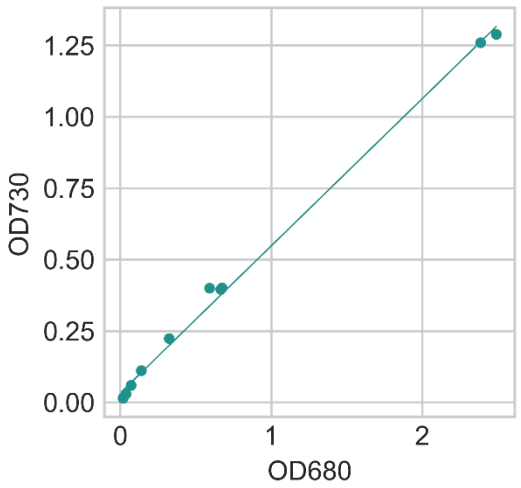
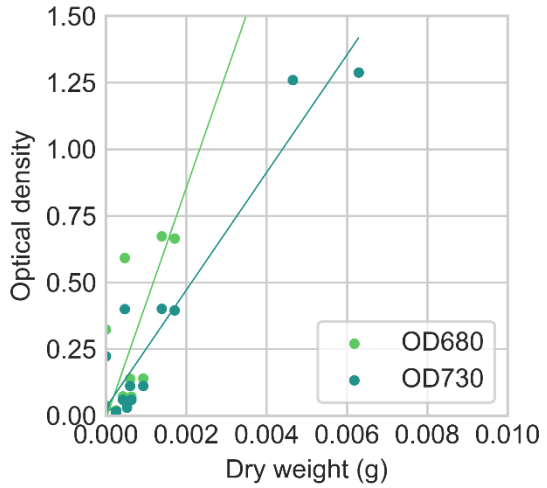


Figure 10: A strong correlation between optical density at 680nm (OD680) and OD730 was found in both A) *Azotobacter vinelandii* DJ ($R^2 \approx 1$) and B) *Synechocystis sp.* PCC 6803 ($R^2 \approx 1$)

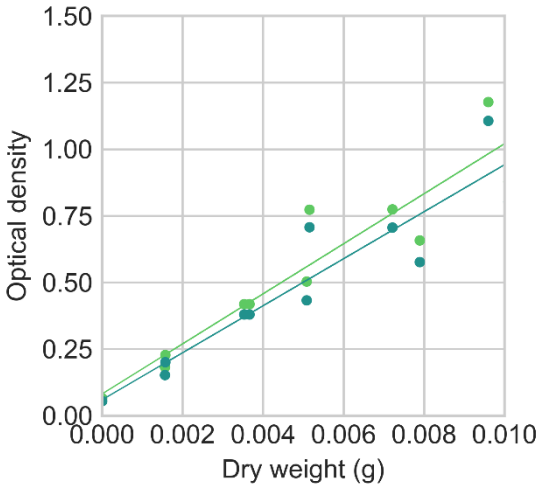
A) *Synechocystis sp.* PCC 6803



$$OD680 = 433.53DW - 0.13$$

$$OD730 = 220.91DW - 0.03$$

B) *Azotobacter vinelandii* DJ



$$OD680 = 93.96DW + 0.08$$

$$OD730 = 88.28DW + 0.06$$

Figure 11: Calibration curves were prepared for A) *Synechocystis sp.* PCC 6803 at optical density of 730nm (OD730) ($R^2 \approx 0.91$), OD680 ($R^2 \approx 0.93$) and B) *Azotobacter vinelandii* DJ OD730 ($R^2 \approx 0.86$) and OD680 ($R^2 \approx 0.88$).

I proposed and supervised experiments carried out by Alexandra Cioffi as an MRes student in our group, to evaluate the use of dialysis tubing to physically separate partners while allowing media to pass through unabated. A pouch is formed from the tubing to enclose one culture and placed into a culture flask containing a different species (Smith & Francis, 2016). The setup has however proven too unwieldy to use, as the pouch would frequently slip open or rip during culture, cells also sometimes leaked out without apparent rupture (Figure 12), and it was difficult to open one side to take a sample without spilling the pouch contents and contaminating cultures.

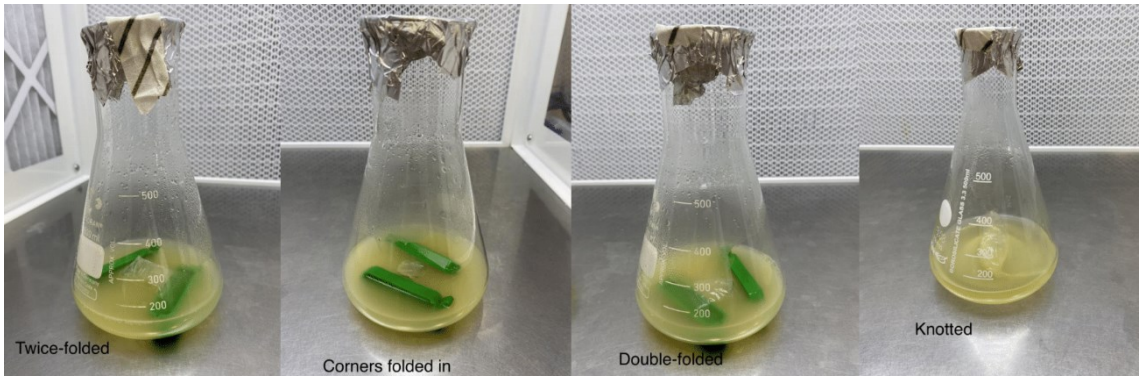


Figure 12: Coculture experiments using *E. coli* in a dialysis pouch in LB medium. Various methods to secure a dialysis pouch were evaluated (twice-folded, corners folded in, double-folded and knotted). Only the bag should contain cells, growth is however apparent in the whole flask, as cells are spilling out. Experiments were carried out and documented by former MRes student Alexandra Cioffi and proposed and supervised by me.

Well inserts (Falcon 353493) – typically used to provide a surface for the adhesion of mammalian cells in multi-well plates with liquid media – were evaluated next. Set up and sampling proved simple with well-inserts. Briefly, a sterile well insert was placed into a well using sterile tweezers, and culture media with or without cells could then be pipetted into the insert or the free space of the well outside the insert (Figure 13). As the insert is permeable to small molecules, media levels equilibrate, and nutrient distribution can be assumed to be equal within the insert and outside of it.

To evaluate the containment of cells, separate experiments were carried out for both species with cells inoculated inside of the insert, as well as outside, using BNxBG11 combination medium. In this set of experiments, the other compartment contained media only. After 3 days, OD measurements were taken from both compartments (Figure 14). Both species were found to be contained within the compartment into which they had been inoculated and not in the other. Notably, the containment is not disrupted by normal shaking along the

horizontal plane during cultivation. There was no significant difference in the growth of *A. vinelandii* when inoculated into insert or underneath it. *Synechocystis* 6803 grew significantly better in the insert than outside ($p < 0.001$), likely as the upper compartment allowed for more photosynthesis, given the mesh at the bottom of the insert is relatively opaque.

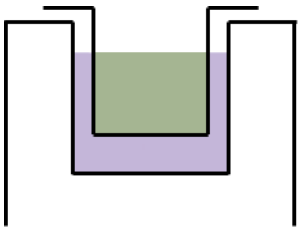
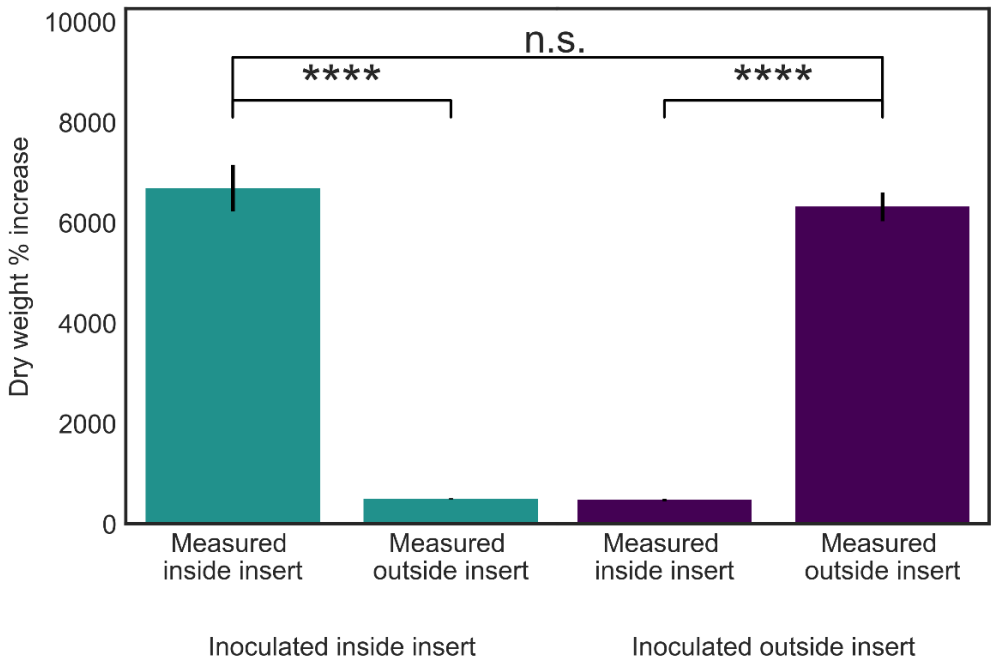


Figure 13: Schematic representation of well insert within a well of 6-well plate. The inside / top culture is designated in green, and the outside / bottom culture is lilac.

A) *Azotobacter vinelandii* DJ



B) *Synechocystis sp.* PCC 6803

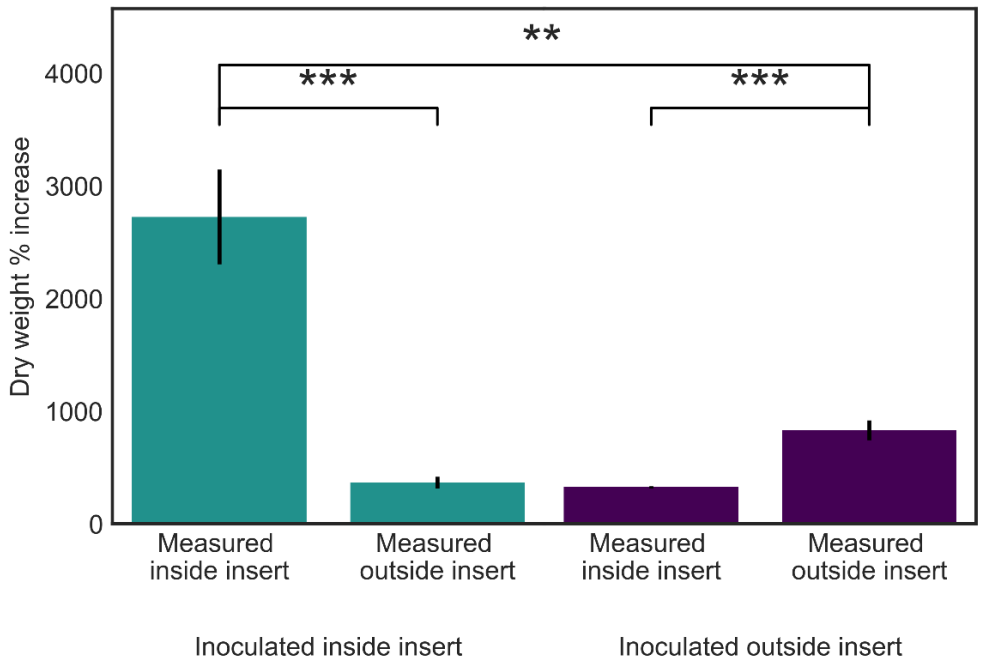


Figure 14: Growth of A) *Azotobacter vinelandii* DJ and B) *Synechocystis sp.* PCC 6803 in 6-well plates with inserts when inoculated inside or outside of a well insert. Stars indicate significance using Student's two-tailed T test, where * = $p > 0.01$, ** = $p > 0.001$, *** = $p > 0.0001$, and **** = $p > 0.00001$. (N = 1, n = 3)

Based on these results, further coculture experiments were set up using 6-well plates with well inserts. *Synechocystis* 6803 was to be grown inside the insert - at the top - and *A. vinelandii* outside the insert - at the bottom - using BNxBG11 combination medium and perturbations thereof with or without additional carbon as 60 mM sucrose and with or without supplemented nitrogen as 20 mM ammonium acetate, as specified earlier in Table 22. To compensate for differential growth rates of the two species, *A. vinelandii* will be inoculated at OD680 = 0.01 and *Synechocystis* 6803 at OD730 = 0.1. Additional work would be required to determine the optimal inoculation ratio.

No difference was found in the growth profiles of *A. vinelandii* (Figure 15) and *Synechocystis* 6803 (Figure 16) in axenic and coculture conditions. Growth curves were fit using the Gompertz equation (Zwietering et al., 1990):

$$y = a \times e^{-e^{b-cx}}$$

This can be reformulated as:

$$y = A \times e^{-e^{\frac{\mu_m e}{A}(\lambda - t)+1}}$$

where:

$$\mu_m = maximal\ growth\ rate$$
$$\lambda = lag\ time$$
$$A = asymptote$$
$$t = time$$

Maximum growth rate (Table 24) was recalculated from the Gompertz fit as:

$$\mu_m = \frac{ac}{e}$$

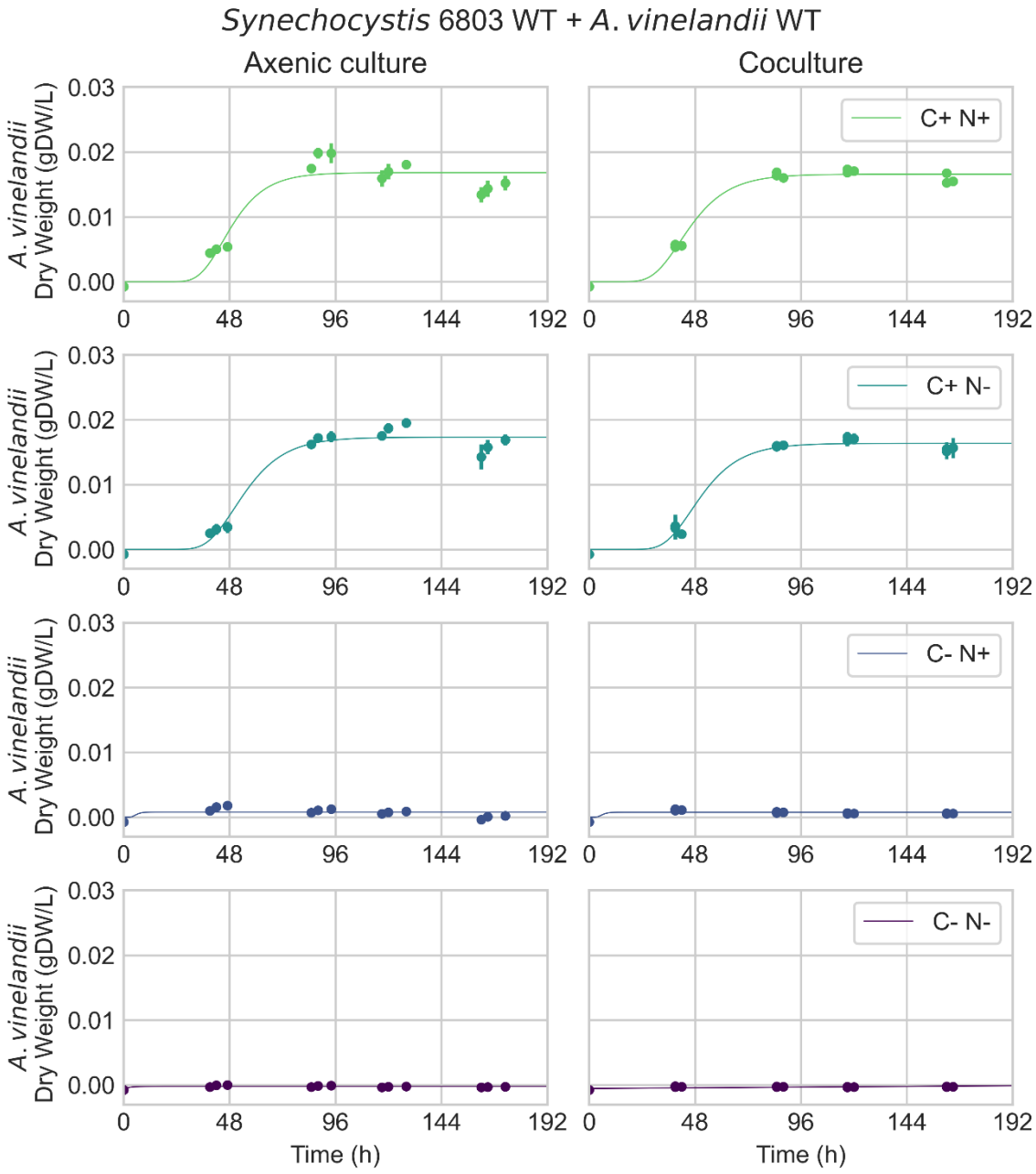


Figure 15: Growth curves of *A. vinelandii* wild-type comparing axenic conditions and coculture with *Synechocystis* 6803 wild-type. Cocultures were grown in BNBG11 combination media perturbations with or without source of carbon as 60 mM sucrose and with or without supplemented nitrogen as 20 mM ammonium acetate. Wild type coculture does not improve growth of *A. vinelandii* without carbon supplementation. Curves fit Gompertz equation (Zwietering et al., 1990). Error bars represent standard deviations for individual experiments (N = 3, n = 9).

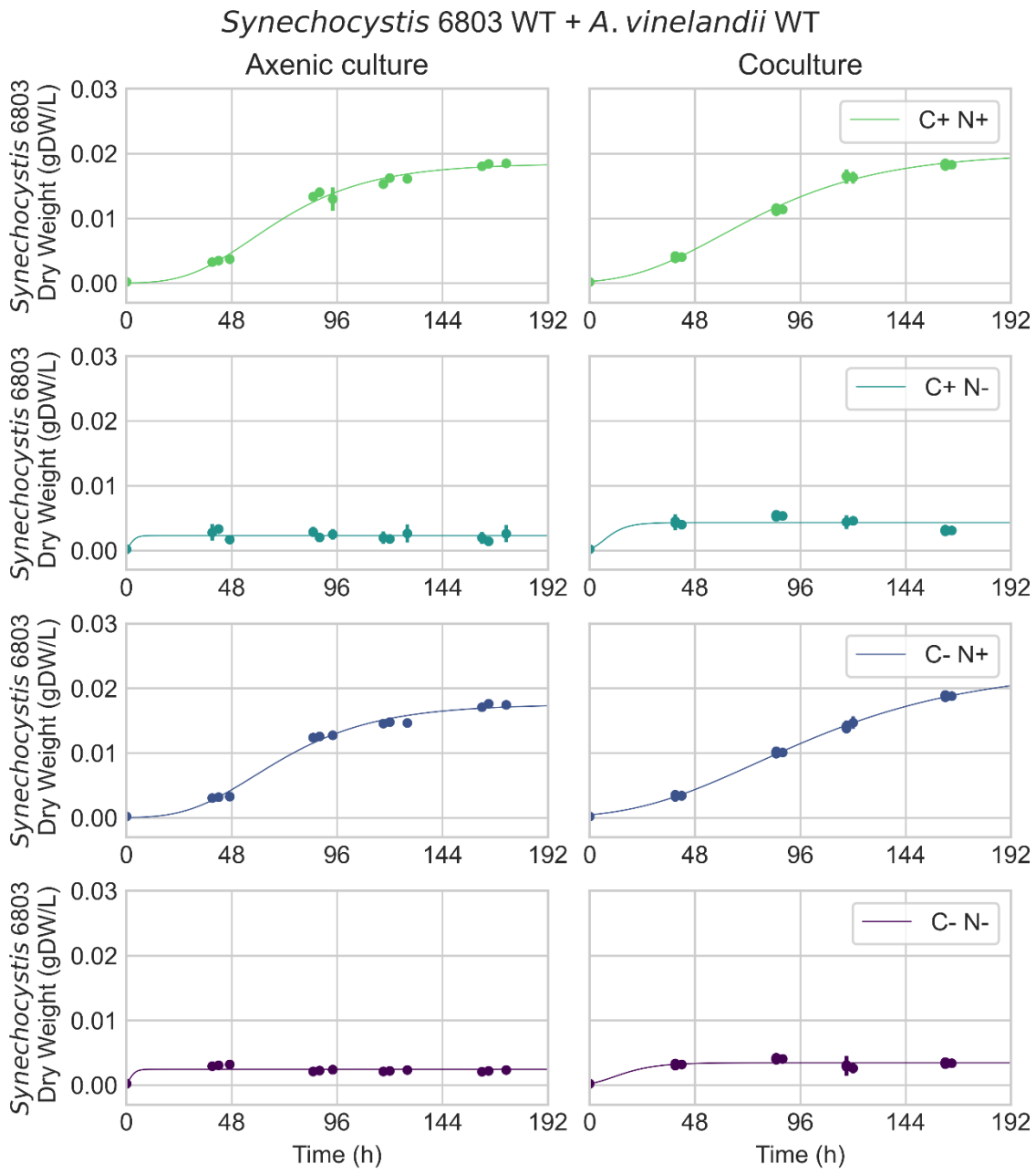


Figure 16: Growth curves of *Synechocystis* 6803 wild-type comparing axenic conditions and coculture with *A. vinelandii* wild-type. Cocultures were grown in BNBG11 combination media perturbations with or without source of carbon as 60 mM sucrose and with or without supplemented nitrogen as 20 mM ammonium acetate. Wild type coculture does not improve growth of *Synechocystis* 6803 without nitrogen supplementation. Curves fit Gompertz equation (Zwietering et al., 1990). Error bars represent standard deviations for individual experiments (N = 3, n = 9).

Rich BNxBG11 was able to support the growth of both partners, while only the diazotrophic partner showed growth in nitrogen-free media, and only the cyanobacterium grew in carbon-free media. A medium lacking both carbon and nitrogen did not support the growth of either axenic culture or of the wild-type consortium, suggesting that any source of fixed carbon or nitrogen produced by either consortium partner is not produced, released, or taken up in sufficient quantities to support the growth of the other partner, and therefore a pathway needs to be engineered instead.

Table 24: Maximum growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 in BNBG11 media perturbations, as calculated from Gompertz fit (see Figure 15, Figure 16). Empty cells indicate no growth.

	A. vinelandii		***Synechocystis* 6803**	
	axenic	coculture	axenic	coculture
rich	5.85×10^{-4}	5.04×10^{-4}	2.29×10^{-4}	1.83×10^{-4}
no N	5.14×10^{-4}	4.99×10^{-4}	-	-
no C	-	-	2.10×10^{-4}	1.53×10^{-4}
no N, no C	-	-	-	-

3.3 Creation of Consortium Model *in silico*

3.3.1 Genome-scale Metabolic Models and Constraint-Based Modelling

Genome-scale metabolic models (GEMs) are constructed from annotated genome sequences and a curated metabolic knowledge base to represent the metabolic capabilities of a given organism. These are represented in the form of a stoichiometric matrix (S matrix), a mathematical representation of the metabolic network, in which rows represent metabolites and columns reactions and entries correspond to the stoichiometry of a reactant in a reaction (Fang et al., 2020). The mathematics underlying the reconstruction and analysis of GEMs have been reviewed by Pinzon et al. (2019).

Flux balance analysis (FBA) is a mathematical tool to calculate the flow of metabolites through a metabolic network under a set of specific constraints (Orth et al., 2010b). The most common tool for FBA is Constraint-Based Reconstruction and Analysis (Cobra) (Price et al., 2004), which uses a set of constraints to model achievable cellular phenotypes using genome-scale biochemical reaction networks, where physico-chemical, topobiological, environmental and regulatory constraints are translated into a set of mathematically-defined balances and bounds, allowing the prediction of the flux distribution under different conditions. Along with stoichiometries, flux bounds represent the allowable flux distributions within a system, i.e. the possible rates at which all metabolites can be produced and consumed. Cobra Toolbox 3.0 (Heirendt et al., 2019), the up-to-date version of Cobra, is freely available as a regularly updated programming toolbox for MATLAB, Python (Ebrahim et al., 2013) and Julia (distributions are available on https://opencobra.github.io/); the MATLAB version offers the widest range of functionality and was therefore the version implemented in this project.

Along with a set of constraints, a specific metabolic state is further defined by its objective function. An objective function is used to calculate the optimal network distribution to achieve maximum flux for the objective function specified. When using FBA to optimise a bioproduction strain, the reaction synthesising a product of interest may be set as the objective function, enabling the observation of changes in metabolism associated with improved flux through this reaction. This can later be used to guide strain engineering. In nature, we commonly work under the assumption that the metabolism of an organism is "configured" in a way to maximise its evolutionary fitness, and hence its rate of growth and cell division; this assumption holds best in wild-type single-cellular organisms (Varma & Palsson, 1994). In GEMs, growth and division is represented by the biomass objective function – a pseudoreaction representing the rate of formation of biomass precursors in their correct proportions (Feist & Palsson, 2010). Thus, by solving FBA with the biomass function as the objective, i.e. the reaction through which flux is to be maximised, we acquire 1) the flux distribution leading to the most possible growth and cell division, and 2) the highest rate of growth itself, in the form of biomass flux.

A biomass objective function is formulated based on the nutrient requirements and composition of a cell, as well as the energy necessary to convert nutrients into macromolecular building blocks. Actual growth rate is constrained by nutrient uptake and maintenance energy requirements. Non-growth maintenance must also be accounted for in the reconstruction a biomass objective function – to grow and divide, the organism must survive in the first place. A biomass objective function remains a useful tool even in the case of modelling the bioproduction of molecules of interest in biotechnological chassis. It is unrealistic to expect that a 100% of chassis resources would be dedicated to the synthesis of the product of interest, neglecting cellular maintenance, growth, or division. Simulations that specify a minimum

required flux through the biomass function while optimising for product formation are sometimes used to model production more realistically.

3.3.2 Implementation of the Two-Cell Model

In this chapter, I developed a combined model of the *Azotobacter-Synechocystis* consortium using the genome-scale model for *Synechocystis* 6803 (iSynCJ816) (Joshi et al., 2017) and *A. vinelandii* (iAA1300) (Alleman et al., 2021). As with all published reconstructions, these models draw on a combination of earlier publications, including models for model species, such as *E. coli*, genomic information, and fluxomics analysis. The steps involved in the reconstruction of genome-scale metabolic models have been described in depth by Monk et al. (2014); Orth et al. (2010a); Thiele & Palsson (2010).

The validity of input models was verified using the Systems Biology Markup Language (SBML) online validation tool (https://sbml.bioquant.uni-heidelberg.de/validator_servlet/). The iSynCJ816 model was found invalid due to a DNA drain reaction. SBML verification was successful once this reaction was removed. Source models were further adjusted to allow flux between -1000 and +1000 mmol/gDWh as the upper infinite and lower infinite bounds, as is standard in FBA reconstructions; irreversible reactions were constrained to 0 mmol/gDWh forward or reverse flux depending on reaction directionality (Becker et al., 2007; Heirendt et al., 2019). For clarification, a flux of -1000 or +1000 mmol/gDWh effectively translates to the reaction being able to carry unlimited flux for the purposes of the model.

To merge the models, first, all reaction and metabolite names, including extracellular metabolites, were modified to include the species name for clarity. Subsequently, struct fields were combined, and the stoichiometric matrix was extended to cover both original matrices. An additional set of diffusion pseudoreactions – denoted with the abbreviation "IEX" (intracellular exchange) in the model – were added for investigated exchange metabolites. The

merged model includes biomass functions for the growth of *A. vinelandii*, as well as for the heterotrophic, phototrophic, and mixotrophic growth of *Synechocystis* 6803 (Table 25). The biomass objective function for autotrophic growth of *Synechocystis* 6803 is the biomass objective function optimised in future references to the *Synechocystis* 6803 biomass objective function.

Table 25: Biomass functions available in the joint consortium model

Reaction name	**Reaction explanation**
Synechocystis_BIOMASS_Ec_SynAuto_1	Autotrophic growth of *Synechocystis* 6803
Azotobacter_BIOMASS_Av_DJ_core	*A. vinelandii* DJ biomass objective function - core
Synechocystis_BIOMASS_Ec_SynMixo_1	Mixotrophic Biomass Growth of *Synechocystis* 6803
Synechocystis_BIOMASS_Ec_SynHetero_1	Heterotrophic Biomass Growth of *Synechocystis* 6803

3.3.3 Media and Other Biologically Relevant Constraints

The model was further constrained to include a simulation of cell-external conditions such as culture media composition and environmental conditions. The composition of BNxBG11 combination media was used as it had previously been experimentally determined as suitable for the culture of *Synechocystis* 6803 and *A. vinelandii* consortium (Section 3.2.1). The concentration of carbon and nitrogen sources in growth media was simulated using the Cobra Toolbox 3.0 `setMediumConstraints()` function (Aurich et al., 2016; Heirendt et al., 2019), which uses cell concentration, cell weight and doubling time to convert the concentration of media components into maximum uptake rates. Maximum uptake rates can

then be used as flux bounds for the appropriate nutrient exchange reaction. Where relevant, media components were separated into ions (Table 26). The import fluxes of micronutrients were left unconstrained.

Table 26: Concentrations (mM) of compounds and ions in media as implemented in setMediumConstrains().

Compound	**Abbreviation**	**Rich media**	**C- N+ media**	**C+ N- media**	**C- N- media**
Sucrose	sucr	58.42828	0	58.42828	0
Ammonium ion	nh4	20.10715	20.10715	0	0
Nitrate (1-) ion	no3	17.94052	17.94052	0	0
Acetate	ac	20.00000	0	20.00000	0
Citrate	cit	0.02143	0	0.02143	0

To approximate cells in the exponential growth phase, cell concentration was set at OD680 or OD730 = 1 for *Azotobacter vinelandii* and *Synechocystis* 6803, respectively. Cells per 1 mL at this OD were determined using haemocytometry using a Neubauer Improved Cell Counting Chamber (0.1 mm, 0.0025 mm^2) under a light microscope (Figure 17).

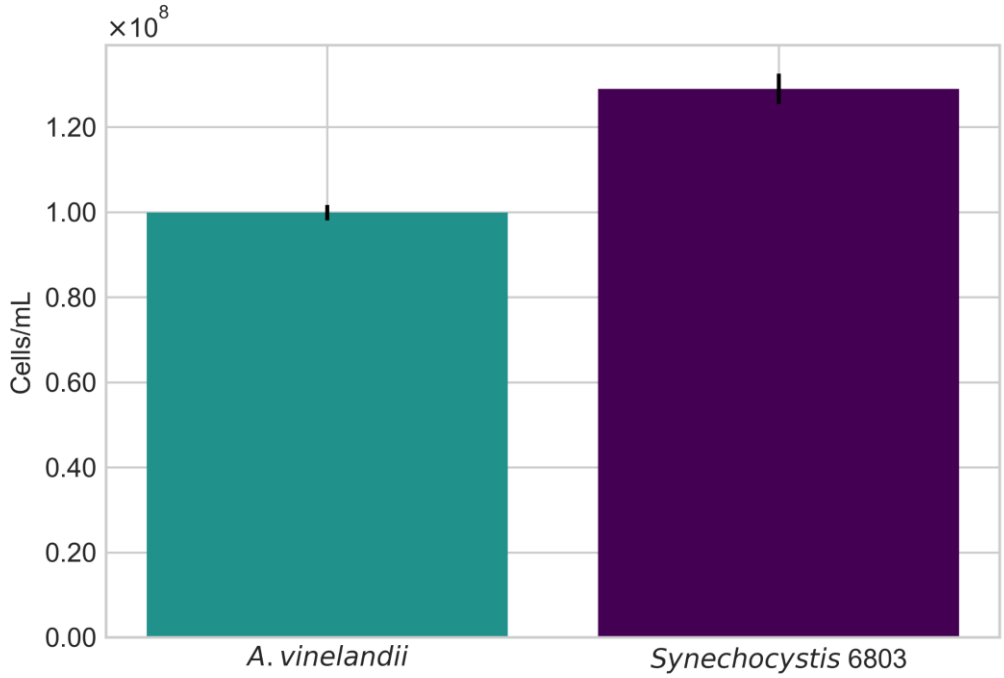


Figure 17: Density of *A. vinelandii* and *Synechocystis* 6803 cells calculated as hundred million (10^8) cells per millilitre based on Neubauer chamber cell counts. Error bars indicate standard error of the mean. Note the y axis is formatted as $\times 10^8$ (hundred million cells/mL). n = 20

A calibration curve correlating OD to dry weight (Figure 11) was used to calculate the weight of a single cell. Cells per 1 mL and weight of a single cell in grams were input into the model rounded to 2 significant figures (Table 27). Doubling times of 3 h and 12 h were used for *A. vinelandii* (Page & Von Tigerstrom, 1979) and *Synechocystis* 6803 (Vermass et al., 1988), respectively.

Table 27: Cell concentration and weight as input into the model based on Figure 17.

	Cell concentration at OD = 1 (cells/mL)	**Weight of single cell (g)**
Azotobacter vinelandii	1.0×10^{8}	1.8×10^{-8}
Synechocystis 6803	1.3×10^{8}	7.1×10^{-9}

N_2, O_2, and CO_2 were also allowed to exchange at maximum exchange rates reported in literature for each species. Photon flux was calculated using an area estimate of 19,245 m^2 per 1 kg *Synechocystis* biomass (Shastri & Morgan, 2005) to mimic 50 $\mu Em^{-2}s^{-1}$, a common illumination setting used in our group when culturing *Synechocystis*. Note that this illumination rate is considerably lower than sunshine illumination on a sunny day. ATP maintenance rate of *A. vinelandii* was constrained to exactly 110 mmol/gDWh to allow a metabolic rate sufficient for respiratory protection of nitrogenase in aerobic conditions (Alleman et al., 2021). Protons, hydrogen gas and water were allowed to exchange freely across the membrane. O_2 uptake and maintenance coefficients chosen for *Azotobacter vinelandii* represent high-aeration conditions (e.g. 10 mL culture in 250 mL flask) consistent with *in vitro* experimental conditions (Castillo et al., 2018). To model photoautotrophic conditions, flux through the oxygen-sensitive *Synechocystis* 6803 hydrogenase was constrained to 0 (Mcintosh et al., 2011) (Table 28).

Table 28: Flux bounds for *Azotobacter vinelandii* and *Synechocystis* 6803 set from literature. (mmol/gDWh)

Organism	**Reaction**	**Lower bound**	**Upper bound**	**Reference**
Azotobacter vinelandii				
	ATP maintenance rate	110	110	(Alleman et al., 2021)
	O_2 uptake	0	83	(Castillo et al., 2018)
	N_2 uptake	0	4.867	(Alleman et al., 2021)
Synechocystis PCC 6803				
	CO_2 uptake	1000	3.7	(Shastri & Morgan, 2005)
	O_2 release and uptake	3.03	0.383	(Joshi et al., 2017) (Yamamoto et al., 2021)
	Photon uptake	0	9.3530787	(Shastri & Morgan, 2005)
	Hydrogenase	0	0	(Mcintosh et al., 2011)

3.4 Model Validation

The ability to sustain positive biomass flux was tested for both models independently in BNBG11-based media with and without a source of fixed carbon and nitrogen in the presence of CO_2 and N_2 gases. While the *Synechocystis* biomass objective function was qualitatively consistent with expected (Table 29) and experimentally observed growth (Figure 15, Figure 16), no optimum solution was found for *A. vinelandii* in this simulated medium.

Table 29: Expected growth profiles in different media perturbations

	***Synechocystis* 6803**	***A. vinelandii***
Rich medium	✓	✓
C-free medium	✓	No growth
N-free medium	No growth	✓
C-free, N-free medium	No growth	No growth

A MATLAB script was used to identify any exchange reactions present in the model prior to the implementation of media constraints whose addition can rescue growth of the constrained model. This resulted in the identification of nickel as an essential metabolite. The essentiality of nickel in the model was further tested by going back to the original model and constraining nickel exchange to zero, which resulted in zero biomass flux. Interestingly, the biomass objective function as defined in the *A. vinelandii* model contains a requirement for nickel. While this is inconsistent with the formulation of Burke's medium, which does not contain any nickel and is routinely used for both diazotrophic and non-diazotrophic *Azotobacter* culture, the inclusion of nickel in growth media has been reported to increase the diazotrophic growth rate of *Azotobacters* (Iswaran & Sundara Rao, 1960), likely due to the presence of a nickel-dependent hydrogenase in *A. vinelandii* that can recycle electrons used in

nitrogen fixation (Menon & Robson, 1994). The inclusion of nickel in the biomass objective function is however likely an artifact of templates used for model generation by (Alleman et al., 2021) going back to an earlier *Azotobacter* model (Tec-Campos et al., 2020) and underlying *E. coli* models (Monk et al., 2017). Models based on earlier publications can lead to error propagation (Monk et al., 2014).

The simulated growth medium was thus reformulated to include $NiCl_2$ (Menon & Robson, 1994) at unconstrained import flux, like all other included micronutrients. This resulted in the predicted pattern of growth / no growth for *A. vinelandii*. No difference was observed in *Synechocystis* 6803 biomass flux with or without nickel. Finally, the model was checked for energy-generating futile cycles, which are thermodynamically infeasible cycles, in which energy molecules (ATP, NADH, NAPH) are charged without the consumption of nutrients. Futile cycles are very common in metabolic models without extensive manual curation (Fritzemeier et al., 2017). This check was carried out by comparing results from FBA with default configuration and loopless FBA; no futile cycles were uncovered.

Maximum predicted growth rates calculated by flux balance analysis (Table 30) were compared with maximum growth rates calculated from *in vitro* data using a Gompertz fit (Table 24). All growth rates exhibit the same qualitative growth / no growth pattern. Predicted growth rates of *Synechocystis* 6803 are in the same order of magnitude (roughly double) as the experimental growth rates. *A. vinelandii* growth rates were predicted to be about 6× higher than measured growth rates.

Table 30: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 in BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells for visual clarity. Compare with Table 24 for *in vitro* data.

	A. vinelandii		***Synechocystis* 6803**	
	axenic	coculture	axenic	coculture
rich	3.45×10^{-3}	3.45×10^{-3}	3.82×10^{-4}	3.82×10^{-4}
no N	3.45×10^{-3}	3.45×10^{-3}	-	-
no C	-	-	3.82×10^{-4}	3.82×10^{-4}
no N, no C	-	-	-	-

3.5 Summary

A BNxBG11 combination medium was developed based on the composition of media typically used to culture both partners individually and confirmed to support the growth of both organisms under investigation. While medium optimisation was not one of the aims of this work, future work could build on this project to enhance the fitness of this coculture, including working with alternative published culture media capable of supporting the growth of the partners. The approach pursued here however demonstrates a simple and practical approach to the development of media for novel cocultures which would be ideally suited to high throughput coculture viability and strain tolerance screens.

I have set up a physical barrier system for the coculture of both species for the separate quantification of the growth profiles of both organisms, while still enabling the exchange of nutrients in consortium. I have not observed the adoption of a well inserts in literature on consortium research, and I believe this user-friendly set-up could greatly benefit researchers in the field, although the high cost of well inserts can limit the throughput of experiments using this set-up. While manufacturer's product specifications specify that the membranes of well inserts are permeable to small molecules, future work would be needed to validate the equilibration of nutrients, especially exchange compounds, across compartments.

Next, I have set up a two-cell model of the consortium compatible with flux balance analysis using Cobra methods (Heirendt et al., 2019) from published genome-scale metabolic models (Alleman et al., 2021; Joshi et al., 2017). This model was modified to enable the treatment of two cell types as separate compartments within the model, with all reactions and metabolites labelled as belonging to one species or the other. Exchange pseudoreactions encoding the exchange of nutrients between cell types were further implemented. The model was finally constrained to reflect the composition of BNxBG11 combination media and the

maximum uptake rates of various nutrients as well as flux through reactions determining a diazotrophic and photoautotrophic phenotype for the partners based on literature evidence.

FBA was used to predict growth rates of *Synechocystis* 6803 and *A. vinelandii* using the constrained GEM. Growth rates of both species were shown to exhibit the same qualitative pattern of growth and no growth relative to the nutrients provided. The model slightly overestimates the growth rates of both species (2× for *Synechocystis* 6803 and 6× for *A. vinelandii*). The predicted growth rates do however sit within the same order of magnitude as the *in vitro* measurements and do not show entirely unrealistic growth rates. For comparison, a maximum photoautotrophic growth rate of 0.085 gDW/h was reported for *Synechocystis* 6803 by Shastri & Morgan (2005), which is roughly 370× higher than the experimental growth rates reported in Table 24.

The difference between *in vitro* and *in silico* measurements can be explained in part by a lack of reaction kinetics and no regulation of gene expression modelled in the GEM. For example, Wu et al. (2019) have reported that the *A. vinelandii* Δ*nifL* mutant, in which feedback repression of nitrogen fixation in response to fixed nitrogen in the cell is removed (discussed further in section 4.3), consumes nearly double the glucose of the wild-type strain during exponential growth to achieve comparable growth rates. It is also possible that maximum aeration could not be achieved using the cultivation set up presented in this chapter, resulting in lower gas exchange than what is modelled, lowering the *in vitro* efficiency of respiration, nitrogen fixation and photosynthesis. Another factor accounting for the discrepancy of *Synechocystis* 6803 growth rates can also be lowered efficiency of photosynthesis, which can be significantly brought down by cell shading; this was not accounted for *in silico*. Finally, the model does not account for dilution of compounds nor for the differential growth rates of the two organisms.

Any discrepancies between model predictions and real growth rates must be born in mind when translating further model predictions to wet lab conditions; in this case, we can expect the model to predict slightly "better" outcomes than may be achieved in the wet lab. This being said, the model predicts growth rates exhibiting the same qualitative patterns and its growth predictions are quantitatively similar to wet lab conditions and remains a useful tool to guide strain engineering, especially when comparing the relative outcomes of alternative pathways, which is precisely its downstream application in this project (Chapter 7).

Chapter 4
Engineering a Nature-Mimicking Nutrient-Exchange System

4.1 Introduction

4.1.1 Rationale

Biological nitrogen fixation is dependent on anaerobic conditions due to nitrogenase's sensitivity to oxygen. Despite this limitation, some organisms have evolved mechanisms to carry out both photosynthesis, which involves oxygen evolution, and nitrogen fixation. The multicellular diazotrophic cyanobacterium *Anabaena sp.* PCC 7120 (further as *Anabaena* 7120) contains two cell types, one of which carries out photosynthesis, and the other carries out nitrogen fixation, being insulated from atmospheric oxygen and photosynthesis. The nutrient exchange between the cell types can be used to guide the design of a consortium, in which two different organisms play the roles of *Anabaena* 7120 two cell types.

The aim of this chapter is to establish a proof-of-concept nutrient exchange: a consortium of the diazotroph *A. vinelandii* and the photoautotroph *Synechocystis* 6803, in which carbon and nitrogen are exchanged analogously to this naturally occurring system. I propose that the set of metabolic pathways required for nutrient exchange in *Anabaena* 7120 can be replicated to establish symbiosis in a diazotroph-photoautotroph consortium, supporting the growth of both cell types without supplemented fixed carbon or nitrogen.

4.1.2 Exchange of Carbon and Nitrogen within the Filamentous Multicellular Cyanobacterium *Anabaena* 7120

Cyanobacteria of the genus *Nostoc* are notable for their ability to fix nitrogen – a strictly anaerobic process – and perform photosynthesis in two types of specialised cells within a

filament, exchanging assimilated nitrogen and carbon backbones across cells. In *Anabaena sp.* PCC 7120 (further as *Anabaena* 7120) – a member of the genus *Nostoc* – photosynthesis is carried out by non-diazotrophic vegetative cells, while nitrogen fixation is compartmentalised in heterocysts. Gases and nutrients passed between cells are exchanged via a combination of a shared periplasm and a microplasmodesmata at the heterocyst-vegetative cell junction (Walsby, 1985). Heterocysts form a protective oxygen-free environment allowing for the functionality of the highly oxygen-sensitive enzyme nitrogenase, which carries out nitrogen fixation. Heterocysts do not express the oxygen-evolving Photosystem II (Tel-Or & Stewart, 1977), and their thick cell walls prevent the entry of oxygen into the cell (Rippka & Stanier, 1978; Walsby, 1985). At the same time, cyclic phosphorylation on the heterocyst Photosystem I serves to provide ATP and electrons as reductant required for nitrogen fixation (Tel-Or & Stewart, 1977). Further adaptations of the *Anabaena 7120* heterocyst are reviewed in Kumar et al. (2010); Malatinszky et al. (2017).

Sources of fixed nitrogen and carbon are exchanged between cells within a filament in what is commonly referred to as the GS/GOGAT cycle (Figure 18). In heterocysts, glutamine synthase (GS) catalyses the amination of glutamate with ammonia synthesised during nitrogen fixation to produce glutamine. This is then transported to vegetative cells, in which it can be broken down alongside α-ketoglutarate (2-oxoglutarate) by glutamate synthase (also known as glutamine:2-oxoglutarate aminotransferase, or GOGAT), recovering glutamate. The amine is then available for protein and amino acid synthesis in the vegetative cell, while glutamate returns to the heterocyst, re-entering the GS/GOGAT C/N exchange cycle. At the same time, the vegetative cells also provide fixed carbon in the form of sugar to both types of cells (Meeks & Elhai, 2002).

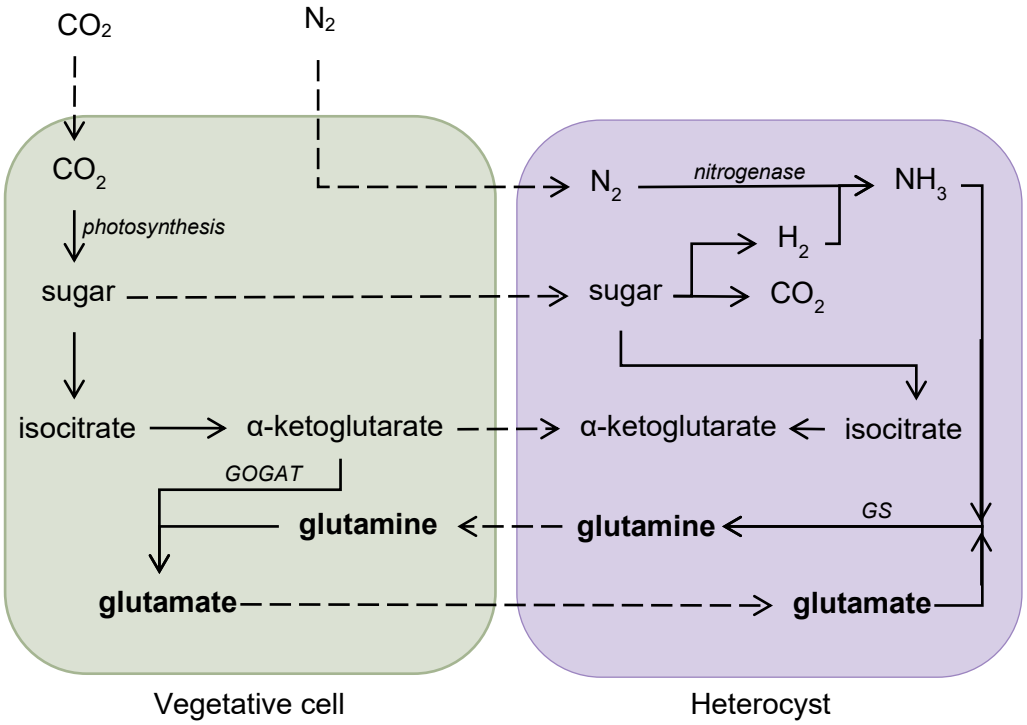


Figure 18: A filament of the diazotrophic cyanobacterium *Anabaena* sp. PCC 7120 showing the flow of nutrients between cells in the GS/GOGAT cycle, where GS is glutamine synthase and GOGAT is glutamate synthase. Adapted from Meeks & Elhai (2002).

4.1.3 Aims

- Design, clone and assemble gene cassettes
- Transform cells
- Develop a method to detect and quantify glutamate and glutamine in media

4.2 Engineering an *Anabaena* 7120-like Nutrient-Exchange System

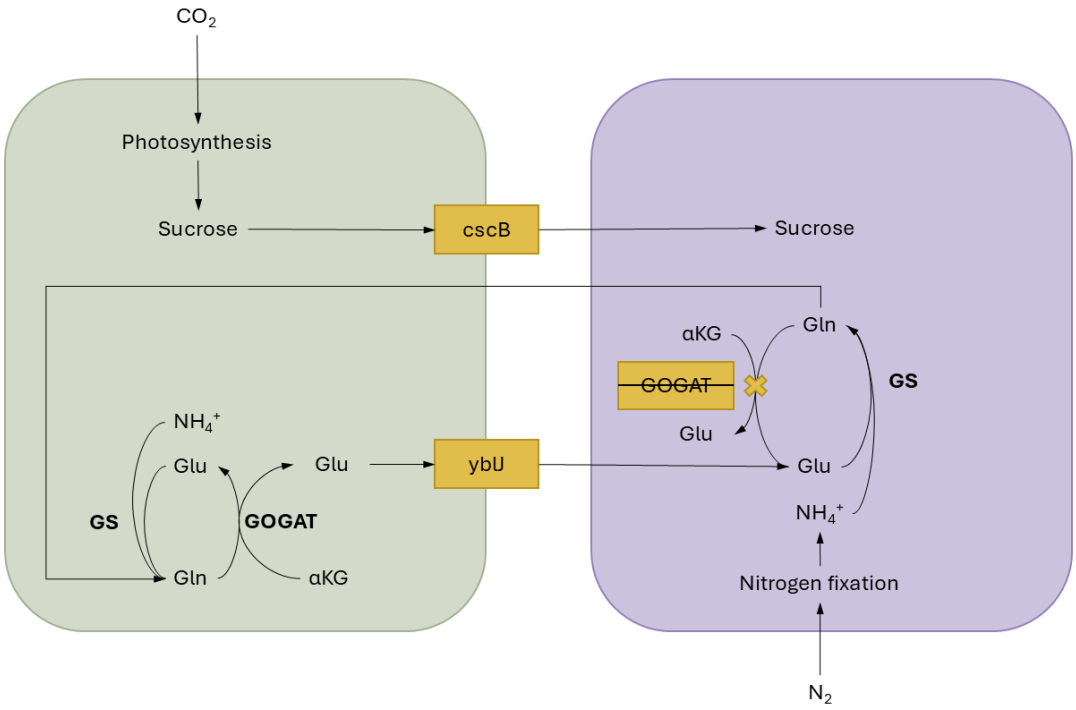


Figure 19: Proposed nutrient exchange mimicking *Anabaena sp.* PCC 7120 using *Synechocystis sp.* PCC 6803 as the photosynthetic partner, and *Azotobacter vinelandii* DJ as the diazotrophic partner, using glutamate (glu) and glutamine (gln) as exchange molecules. A glutamate exporter (yblJ) and sugar exporter (cscB) are introduced into *Synechocystis sp.* PCC 6803 and GOGAT (glutamate synthase) is deleted in *Azotobacter vinelandii* DJ. Water and cofactors are omitted for simplicity.

The metabolic exchange in a nutrient exchange consortium of *A. vinelandii* and *Synechocystis* 6803 which mimics the GS/GOGAT cycle in *Anabaena* 7120 is outlined in Figure 19 and Table 31. First, to improve the availability of fixed nitrogen in *A. vinelandii*, I have introduced a deletion of the feedback regulator *nifL* (Bali et al., 1992). NifL is an anti-

activator protein of the oxygen-sensor NifA, where NifA acts as an activator of transcription of *nif* genes, activating nitrogen-fixation. In the presence of fixed nitrogen, NifA forms an inactive complex with NifL. By deleting NifL, NifA is permanently derepressed, thus removing feedback regulation of nitrogen-fixation (Dixon & Kahn, 2004). The amount of ammonia produced is therefore increased relative to wild-type, as fixed nitrogen present in the cell does not repress further nitrogen fixation.

Native biosynthesis of glutamate in *A. vinelandii* was disrupted through the deletion of glutamate synthase (GOGAT) to induce a dependence on glutamate released by *Synechocystis* 6803. The introduction of glutamate exporter yblJ (Matsudaira et al., 2020) into *Synechocystis* 6803 supplements its passive release of glutamate into media. Free glutamate is then taken up by *A. vinelandii* from the medium using a native high-affinity glutamate transporter (Mishra et al., 1991; Smith & Francis, 2016). To provide an additional source of fixed carbon for the diazotroph in addition to glutamate only, the cscB sucrose exporter was to be introduced into *Synechocystis* 6803 (Smith & Francis, 2016).

Table 31: Genetic modifications of *A. vinelandii* and *Synechocystis* 6803 to induce a nature-mimicking exchange

Background	**Modification**	**Purpose of modification**
Azotobacter vinelandii DJ	Δ*nifL*	Overproduction of fixed nitrogen
	Δ*GOGAT*	Disruption of glutamate biosynthesis
Synechocystis sp. PCC 6803	*cscB*	Sucrose export
	yblJ	Glutamate export

To create gene deletions in *A. vinelandii*, I used double reciprocal homologous recombination (HR)-mediated replacement with an antibiotic cassette (Figure 20). *A. vinelandii* is considered highly recombinogenic, making it suitable for HR-based genetic techniques (Dos Santos, 2011). I have designed homology arms of 150 bp homologous regions up- and downstream of the target gene. A knockout cassette was built using BASIC assembly, where an antibiotic resistance cassette is flanked by upstream and downstream HR arms and stored in a storage vector carrying a different antibiotic until ready to be transformed into *A. vinelandii.*

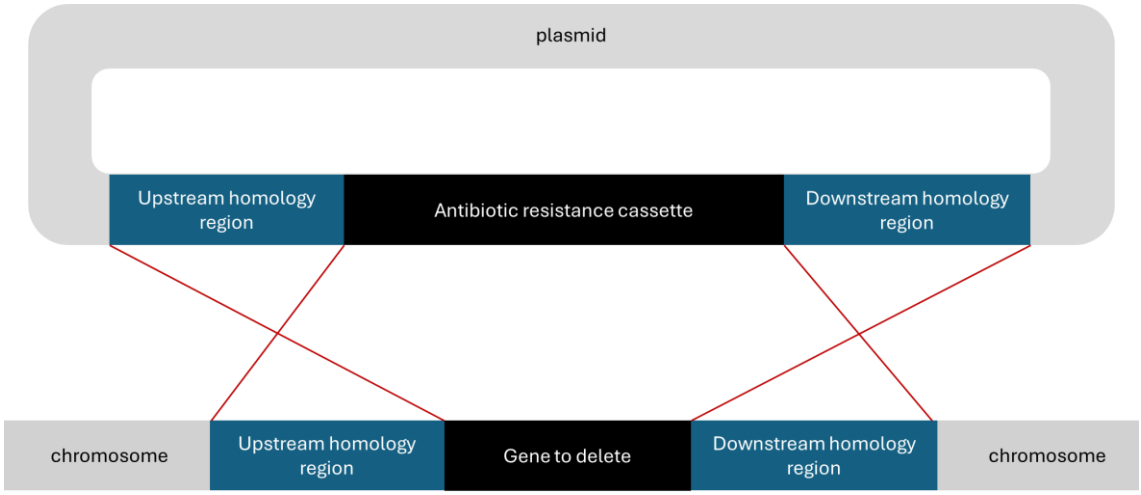


Figure 20: Schematic representation of homologous recombination-mediated gene deletion.

4.2.1 Predicted Growth Rates

A genome-scale metabolic model with flux-balance analysis (model prepared in Section 3.3) was used to predict the growth rates of *A. vinelandii* and *Synechocystis* 6803 when engineered to mimic the *Anabaena* 7120 nutrient exchange. For clarity, this model was constructed by constraining the flux through the *A. vinelandii* GOGAT reaction to zero to simulate the knockout condition and enabling unconstrained bidirectional transport of sucrose and glutamate between cells. The *nifL* deletion was not modelled, as genome-scale metabolic

models do not account for gene regulation; note that the model itself is however closer to a Δ*nifL* condition than to true wild-type, as deleting this gene alleviates the repression that the model fails to account for in the first place.

This simulation predicts an order of magnitude improvement of growth rate of *A. vinelandii* and two orders of magnitude improvement of growth rate of *Synechocystis* 6803 in the rich (N and C supplemented) media relative to wild-type (Table 32). The same trend is observed for *A. vinelandii* in nitrogen-free, carbon-rich media, and for *Synechocystis* 6803 in nitrogen-rich, carbon-free media – both cells are predicted to achieve the same maximum growth rate in the medium missing the nutrient they can synthesise from the atmosphere. Notably, the prediction suggests that both cell types can achieve maximum growth rates comparable to those in rich media when cultivated in media lacking their respective essential nutrient, and in media lacking both fixed carbon and nitrogen if this nutrient exchange is introduced; this is consistent with the survival of *Anabaena* 7120 in conditions deficient in both fixed carbon and fixed nitrogen.

Finally, it must be highlighted that the predicted maximum growth rates reported below are indeed *maxima* and do not take into consideration the maintenance demands of the "donor" strain, or the strain not being optimised for in each condition. The optimisation process instead assumes that the donor is solely dedicated to supporting its partner, with zero flux through its own biomass objective function.

Table 32: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium engineered to mimic the nutrient exchange in *Anabaena* 7120 per Figure 19 and Table 31. Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	mimic	wild-type	mimic
rich	3.45×10^{-3}	3.02×10^{-2}	3.82×10^{-4}	2.20×10^{-2}
no N	3.45×10^{-3}	3.02×10^{-2}	-	2.19×10^{-2}
no C	-	2.80×10^{-2}	3.82×10^{-4}	2.20×10^{-2}
no N, no C	-	2.80×10^{-2}	-	2.19×10^{-2}

4.2.2 Cloning and Amplification of Bioparts

Bioparts Assembly for Idempotent Cloning (BASIC) (Storch et al., 2015) was chosen as the assembly method for this project; all parts were therefore created or purchased in BASIC standard format. BASIC defines standard bioparts using standardised prefix and suffix regions, which, when digested with type II restriction enzyme BsaI produce sticky ends compatible with a set of linker parts. Directional multipart assembly is enabled through a set of multiple orthogonal linkers, which define the order of bioparts within a construct. An additional set of methylated linkers within the BASIC toolkit also enables hierarchical assembly, as methylated linkers contain an additional BsaI site protected by methylation. The protected site does not get digested during first use, but methylation is eventually lost during replication in chassis, making the site accessible for digestion in downstream cloning. This cloning strategy was also

chosen as it was already established in our lab. Sequences of BASIC linkers used in this study are available in Supplementary table 2.

To create BASIC-compatible parts, internal BsaI sites were manually removed from sequences in the Benchling web application (Benchling, 2021). All parts were flanked the upstream prefix 5'-TCTGGTGGGTCTCTGTCC-3' and downstream suffix 5'-GGCTCGGGAGACCTATCG-3'. Parts were codon optimised for the recipient organism using the IDT codon optimisation tool (https://eu.idtdna.com/CodonOpt). The sequences of homology arms, which necessitate high sequence identity, were not codon optimised.

Parts were purchased from IDT as gBlocks, or when GC content was too high for synthesis as gBlocks, they were purchased as clonal parts instead. Clonal parts were immediately transformed into *E. coli* DH5α and stored as glycerol stocks, and gBlocks were subcloned into pJET1.2 using blunt-end cloning with the CloneJET kit (Thermo Scientific) first following manufacturer's instructions, and later using a modified protocol. The subcloned part was then transformed into competent *E. coli* DH5α cells and plated onto an antibiotic-selective plate.

The pJET1.2 plasmid contains dual selection, an antibiotic resistance cassette within the backbone to confirm successful transformation, and a lethal gene, which gets interrupted and becomes non-functional upon successful ligation of insert; only colonies transformed with a plasmid into which the insert has been successfully ligated should therefore be viable on a plate containing appropriate antibiotic. As a negative control, the heat-shock transformation protocol was also carried out with competent cells without any plasmid. Colonies were observed on all true subcloning plates and were screened by colony PCR using primers annealing near the insertion site on the pJET1.2 backbone. All colonies were also kept re-streaked onto fresh plates for future reference.

Table 33: DNA bioparts ordered for synthesis for nature-mimicking nutrient exchange. All sequences are listed in Supplementary table 1.

Shortcode	**Description**	**Target strain**	**Synthesis method**
AV_nifL_down_BASIC	nifL downstream region	*Azotobacter vinelandii DJ*	gBlock
AV_nifL_up_BASIC	nifL upstream region	*Azotobacter vinelandii DJ*	gBlock
ybjL_BASIC	ybjL glutamate transporter from *E. coli*	*Synechocystis* 6803	gBlock
cscB_BASIC	*cscB* sucrose-proton symporter	*Synechocystis* 6803	gBlock
AV_GOGAT_up_BASIC	GOGAT upstream region	*Azotobacter vinelandii DJ*	Clonal part
AV_GOGAT_down_BASIC	GOGAT downstream region (NOTE: 1bp change to avoid BsaI site)	*Azotobacter vinelandii DJ*	Clonal part

The efficiency of subcloning using the CloneJET kit was found to be extremely low, producing only 3 verified colonies out of a set of 319 tested across several rounds of cloning, and only one of four gBlocks (*A. vinelandii* nifL upstream region: AV_nifL_down_BASIC) was successfully cloned using this strategy. This part was verified by Sanger sequencing. While the CloneJET protocol specifies that a considerable amount of recircularised DNA remains on

the plate after transformation, which can give false negatives in colony PCR, some true positives should still be expected.

When no bands of the required length were produced by colony PCR for the majority of bioparts after three rounds of cloning, I tried improving my colony PCR method. Cell lysis conditions were first modified from a 5 min at 95°C step in the thermocycler at the start of the PCR protocol itself to alternatingly placing the tube in a -80°C freezer and 95°C water bath (technique suggested by Dr Tom Oliver), followed by centrifugation to remove the pellet. PCR primers provided with the kit were also replaced by similar primers previously designed and routinely used by Jonathan Muller (fellow PhD student in Jones group), testing both old stocks and newly synthesised primers. A gradient of annealing temperatures (50C°, 52°C, 56°C and 60°C) was also tested. Bands of the expected length were not acquired in any condition tested. Only very short, smudged bands (likely corresponding to primer dimers) were observed. Miniprepped plasmid DNA was also submitted for Sanger sequencing, in case faulty colony PCR was showing false negatives, but no reads were produced. Not even reads corresponding to the empty backbone were seen, even with good miniprep yield (yield ≈ 100 ng/µL, ratio < 2).

Next, I trialled overlap-extension PCR (OE-PCR) based cloning (Bryksin & Matsumura, 2010). Primers were designed to anneal to BASIC standard suffix and prefix, with an overhang homologous to pJET1.2 regions flanking the insertion sequence; after one round of amplification, the new part was used as a megaprimer in the second round of PCR to insert into pJET1.2. Product was then transformed into *E. coli* DH5α using heatshock. As with the CloneJET strategy, colonies were observed on selective media, but the presence of insert could not be confirmed after colony PCR.

Given that cell growth on selective media was observed in all set-ups, but not in negative controls containing untransformed cells only, it appears that the killing cassette included in pJET1.2 was not lethal. Only positive selection for transformation (by proxy of antibiotic resistance) was observable, regardless of the presence of insert. The other member of my group also confirmed observing leaky activity, where most colonies he produced after CloneJET did not contain insert, although at least one positive clone was usually produced.

To check whether cells carrying a self-ligated plasmid survived, I carried out the subcloning protocol with pJET1.2 plasmid only and no insert, followed by a transformation. This did result in colonies, suggesting that cells were indeed surviving without insert. For this reason, I decided to avoid the pJET1.2 plasmid in further cloning and explore alternatives.

I proceeded to subclone the bioparts using BASIC assembly directly, flanking with methylated linkers LMP and LMS for idempotency (Storch et al., 2015). Remaining gBlocks were individually assembled into an kanamycin antibiotic-selective GFP-dropout backbone available in the Jones group (pJM01, Table 34) per Figure 21, transformed into competent *E. coli* DH5α using heatshock transformation, and plated onto antibiotic selective plates. This cloning method also contains dual selection. All successfully transformed colonies are viable on kanamycin LB plates. Presence of insert is validated by lack of green glow of colonies under UV light, as insert has replaced GFP dropout cassette.

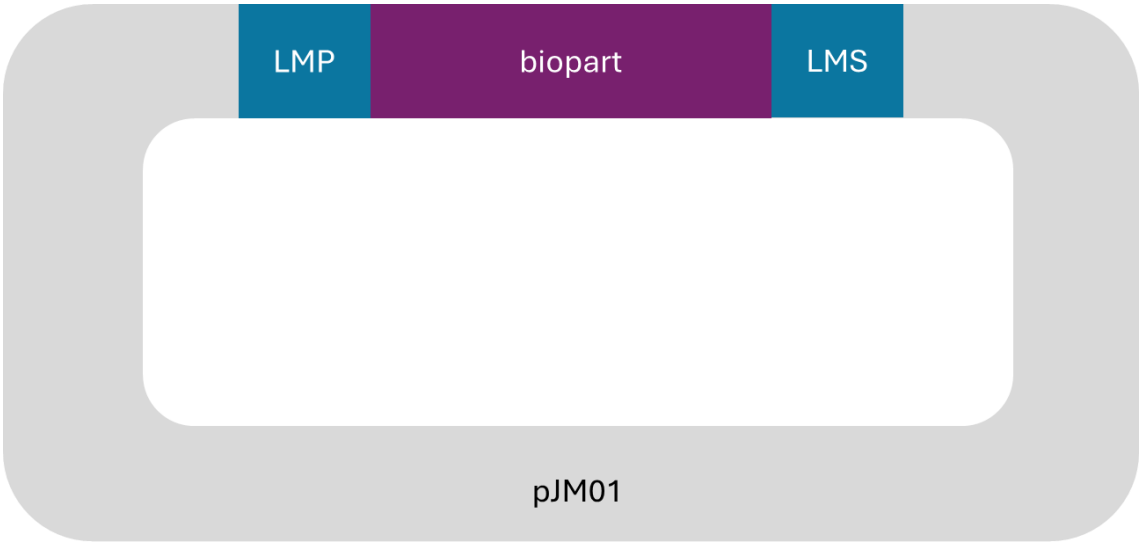


Figure 21: Schematic representation of biopart stored using BASIC assembly with methylated linkers (LMP and LMS)

DNA from non-glowing colonies was subsequently isolated using minipreps, followed by sequence validation. This yielded successfully cloned *A. vinelandii* nifL downstream region (AV_nifL_down_BASIC) and *cscB* sucrose-proton symporter (cscB_BASIC). Only ybjL glutamate transporter from *E. coli* (ybjL_BASIC) was not cloned, as this method could not be repeated as all wet lab work had to be halted due to sudden onset of a medical condition. Cloned parts are summarised in Table 35.

Table 34: Plasmid backbones used in this work

Plasmid name	Description	Resistance	Source
pJET1.2	pMB1-eco47IR-PlacUV5-T7-AmpR	Ampicillin	(Thermo Fisher Scientific Inc., 2006-2024)
pJM01	pET-lac-GFP-KanR	Kanamycin	Jonathan Muller
pJM02	pCDF-lac-GFP-SpecR	Spectinomycin	Jonathan Muller
pJM03	pACYC-lac-GFP-GmR	Gentamicin	Jonathan Muller

Table 35: List of DNA parts and subcloning strategies

Shortcode	Description	Storage backbone	Successful cloning strategy
AV_nifL_down_BASIC	nifL downstream region	pJM01	BASIC
AV_nifL_up_BASIC	nifL upstream region	pJET1.2	CloneJET
ybjL_BASIC	ybjL glutamate transporter from *E. coli*	none	none
cscB_BASIC	*cscB* sucrose-proton symporter	pJM01	BASIC

4.3 Construction of Δ*GOGAT* and Δ*nifL A. vinelandii* strains

4.3.1 Cloning a *GOGAT* and *nifL* Knockout Cassette

I have engineered a red fluorescent protein expression cassette to simplify visual readout of successful transformation and expression in *A. vinelandii*. Red was chosen to be visually distinctive from the green fluorescence naturally exhibited by *A. vinelandii* under iron starvation (Dos Santos, 2011).

A BASIC-compatible *mCherry* gene was kindly provided by Dr Daniel Zabala Álvarez in our research group. The *mCherry* expression cassette was assembled as follows, from 5' to 3': constitutive BBa_J23119 promoter (https://parts.igem.org/Part:BBa_J23119; Anderson, 2006), followed by medium strength ribosome binding site on BASIC linker (RBS2), *mCherry* gene, neutral linker T3, *E. coli* rrnB T1 terminator (https://parts.igem.org/Part:BBa_B0010, Rettberg, 2003) (Orosz et al., 1991). While *A. vinelandii* is considered a model nitrogen fixing organism with established molecular biology techniques (Dos Santos, 2011), it has not been explored much as a synthetic biology chassis, and therefore the availability of bioparts, and the compatibility of extant characterised bioparts with *A. vinelandii* are yet to be described. Parts used in this work were chosen primarily for their availability in the group's collection as BASIC-formatted bioparts (Yunus, 2019).

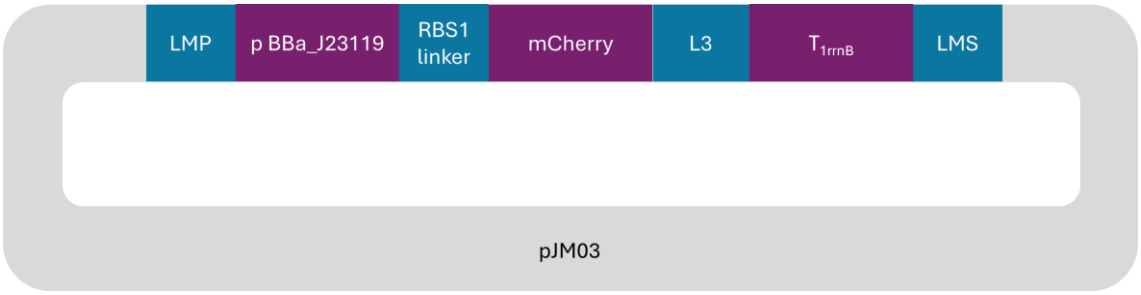


Figure 22: Schematic representation of mCherry cassette carrying BBa_J23119 promoter (p BBa_J23119), RBS1 linker, *mCherry* red fluorescent protein gene, neutral linker 3 and rrnB T1 terminator, all flanked by methylated linkers (LMP and LMS) in a pJM03 backbone (pACYC-lac-GFP-GmR).

A knockout cassette was built using BASIC assembly, where an antibiotic resistance cassette is flanked by upstream and downstream HR arms and stored in a storage vector carrying a different antibiotic (Figure 24) until ready to be transformed into *A. vinelandii*. A kanamycin and ampicillin resistance cassette was assembled analogously to the mCherry cassette outlined above.

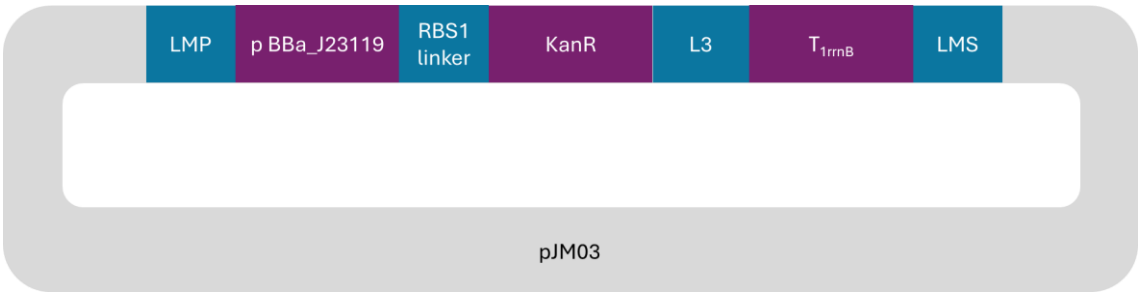


Figure 23: Schematic representation of antibiotic resistance cassette carrying BBa_J23119 promoter (p BBa_J23119), RBS1 linker, kanamycin resistance gene, neutral linker 3 and rrnB T1 terminator, all flanked by methylated linkers (LMP and LMS) in a pJM03 backbone (pACYC-lac-GFP-GmR)

Bioparts containing sequences homologous to regions up- and downstream of the gene in question were assembled as flanking the previously constructed antibiotic resistance cassette or *mCherry* expression cassette, linked by neutral linkers. As the reuse of DNA sequences is associated with undesirable recombination events, unique linkers were used at each level of assembly. The whole resulting knockout cassette is flanked by methylated linkers LMP and LMS and stored in pJM02 storage plasmid (Table 34).

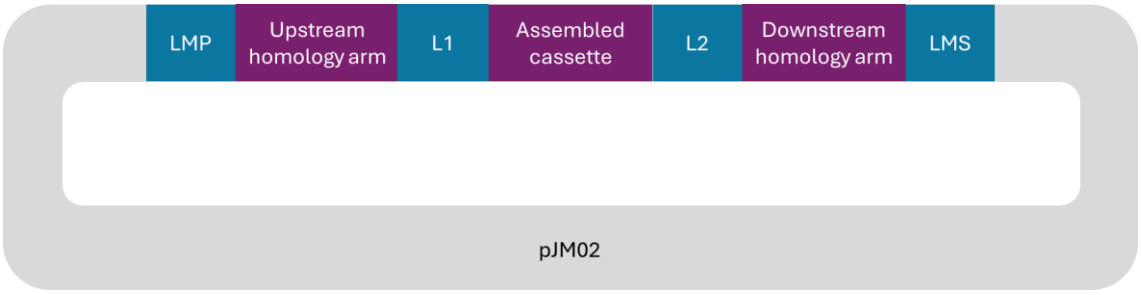


Figure 24: Schematic representation of knockout cassette in storage vector, where parts are connected by neutral linkers (L1 and L2), and surrounded by methylated linkers (LMP and LMS). Assembled cassette refers to either the mCherry cassette (Figure 22) or antibiotic resistance cassette (Figure 23) constructed earlier.

For antibiotic-resistance cassette containing constructs, transformed *E. coli* DH5α were first plated onto a spectinomycin selective plate, as spectinomycin resistance is conferred by the plasmid backbone, followed by a replating of non-luminescent colonies onto a kanamycin selective plate. The same approach was used with the ampicillin resistance cassette used in the nifL deletion cassette. *E. coli* DH5α transformed with the deletion cassette storage plasmids in question were found to survive on both spectinomycin plates, and on the antibiotic associated with the cassette included – kanamycin for GOGAT, and ampicillin in the case of nifL. Pink colonies were also observed on plates transformed with *mCherry*. Gel electrophoresis after plasmid extraction was used to confirm correct size of assembled plasmid.

4.3.2 Transformation of *A. vinelandii*

A. vinelandii was transformed using the iron-starvation method. Briefly, cells precultured in Burke's media were transferred into a modified version of Burke's media supplemented with ammonium acetate (BN) lacking molybdenum and iron until culture turns bright green, signifying the release of siderophore associated with competency (Dos Santos, 2011). Competent cells were subsequently transformed natural transformation (Methods 2.2.3.1).

Transformed cells were plated onto BN plates with spectinomycin, where spectinomycin is the antibiotic on the plasmid backbone in all cases. Pink colonies from plates with mCherry constructs were picked and restreaked for easier visualisation (Figure 25). Colonies from plates with antibiotic resistance constructs were restreaked onto plates with the appropriate second antibiotic only – ampicillin for ΔnifL and kanamycin for ΔGOGAT, respectively. Unfortunately, validation of genomic integration of constructs could not be carried out due to sudden onset of medical condition preventing further wet-lab work, and appropriate steps are discussed in chapter summary (Section 4.5).

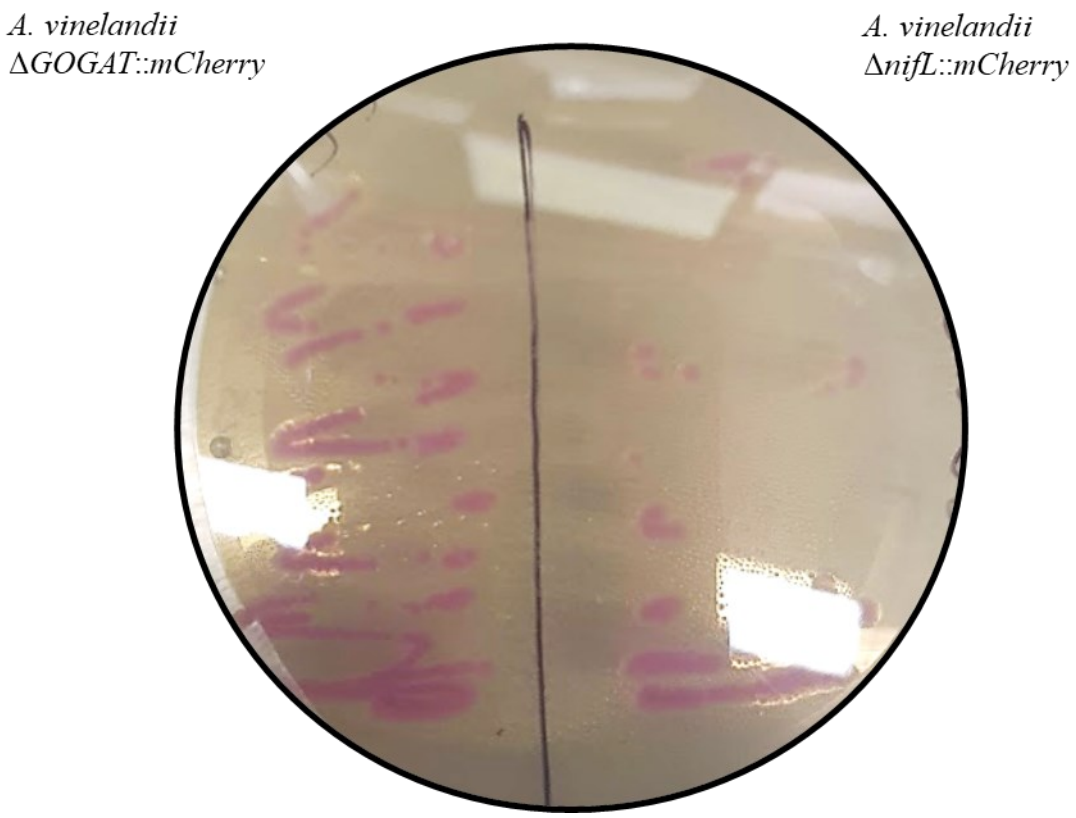


Figure 25: *A. vinelandii* transformed with GOGAT (Δ*GOGAT::mCherry*) and nifL (Δ*nifL::mCherry*) deletion cassettes carrying mCherry showing a clear pink coloration indicating successful transformation. Plated on antibiotic-free BN agar.

4.4 Developing a Method for the Separation and Detection of Glutamate and Glutamine

Aqueous solutions of glutamate and glutamine, and a mixture of both glutamine plus glutamate were prepared for analysis. Ten concentrations between 0 g/L and 1 g/L were tested. Samples were derivatised on-line using o-phthaldehyde (OPA) with sodium sulphite (Material and Methods 2.2.6.1) and separated on an Agilent 1200 series high-performance liquid chromatography (HPLC) system using on a Zorbax Eclipse C18 column (Agilent) with diode-array detection (DAD) at 338 nm and 210 nm (Materials and Methods 2.2.6.2) (Henderson & Brooks, 2010; Henderson et al., 2000; Shafaei et al., 2017; Wang et al., 2010). HPLC peaks corresponding to both underivatized (Figure 26, Figure 27) and OPA-derivatised (Figure 28, Figure 29) glutamate and glutamine at both wavelengths were compared to identify unique peaks that can be used to discriminate between the two amino acids when in mixture.

A unique peak for derivatised glutamate was identified at both 210 nm and 338 nm, retention time (RT) ≈ 1.5 min. A unique peak for derivatised glutamine was found at 338 nm, RT ≈ 2 min. No peaks were detected in underivatized samples at 338 nm. Two peaks were observed in a mixture of underivatized glutamine and glutamate at 210 nm, the resolution of peaks was improved at 338 nm.

Derivatised sample detected at 338 nm

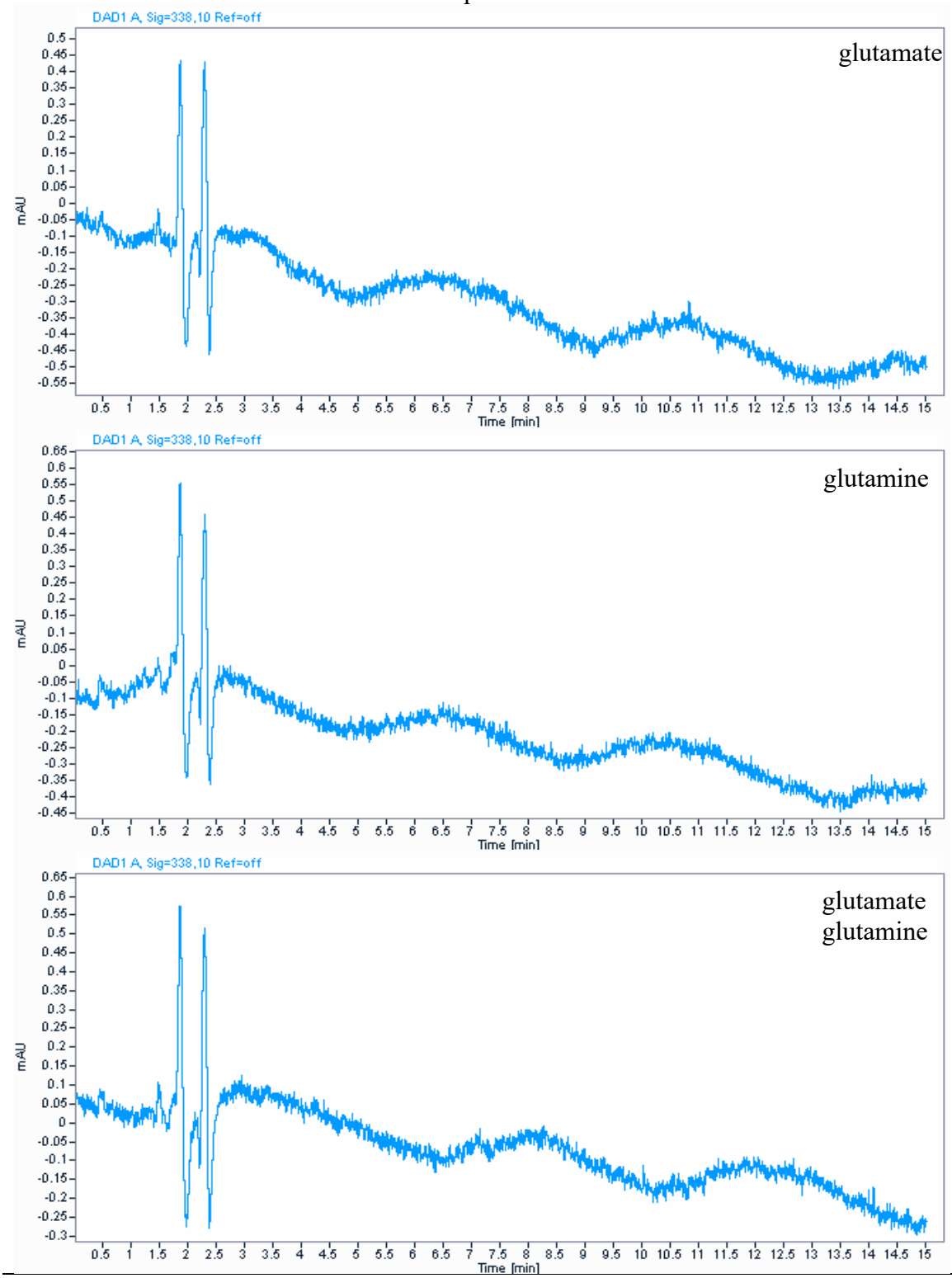


Figure 26: No clear peaks are apparent of HPLC chromatographs of glutamate, glutamine or a mixture of the two when detected at 338 nm without derivatisation. All samples at 0.5 g/L concentration. n =3

Underivatised sample detected at 210 nm

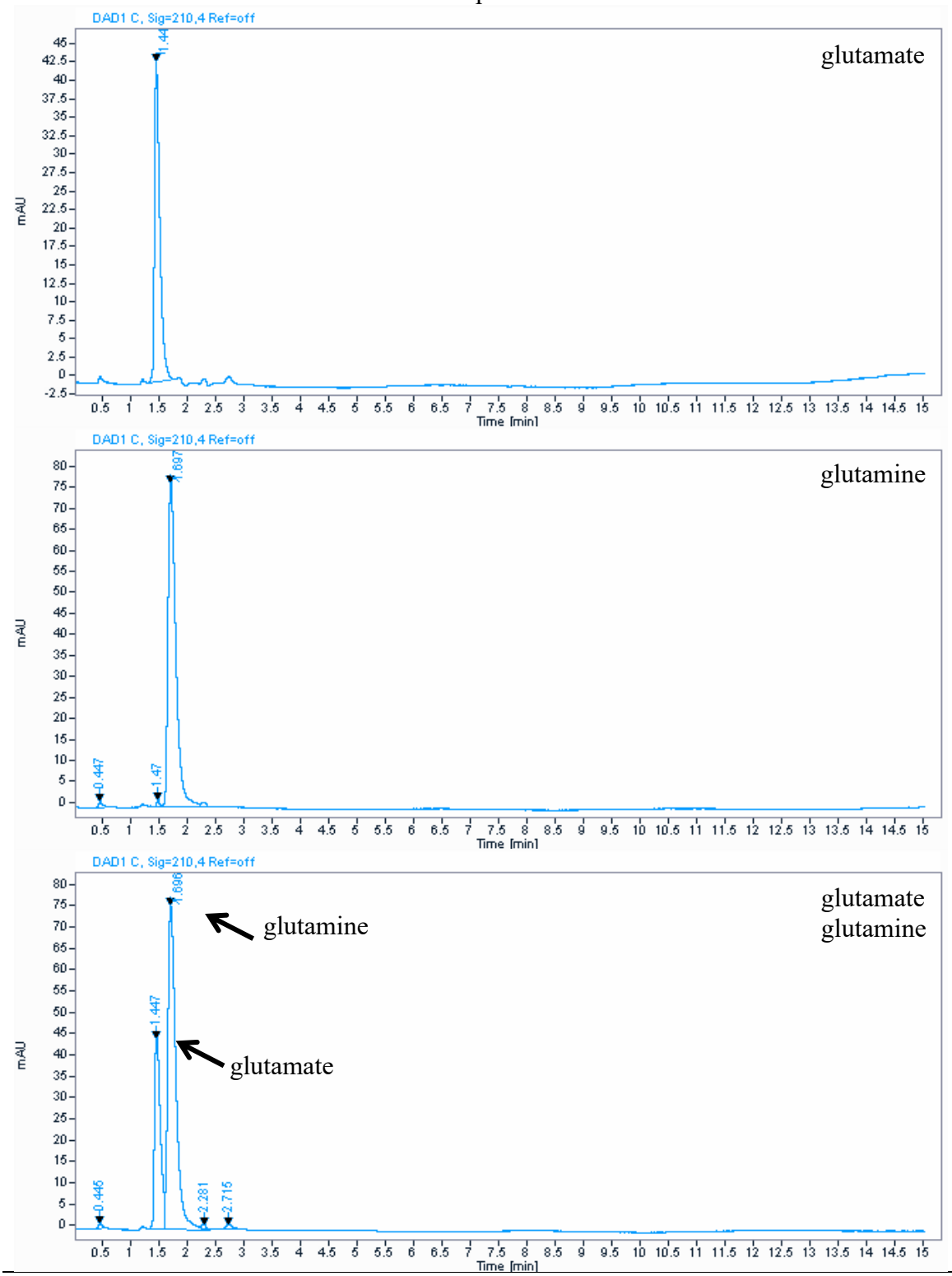


Figure 27: HPLC chromatographs for underivatised glutamate and glutamine at 210 nm show distinct peaks. All samples at 0.5 g/L concentration. n =3

Derivatised sample detected at 210 nm

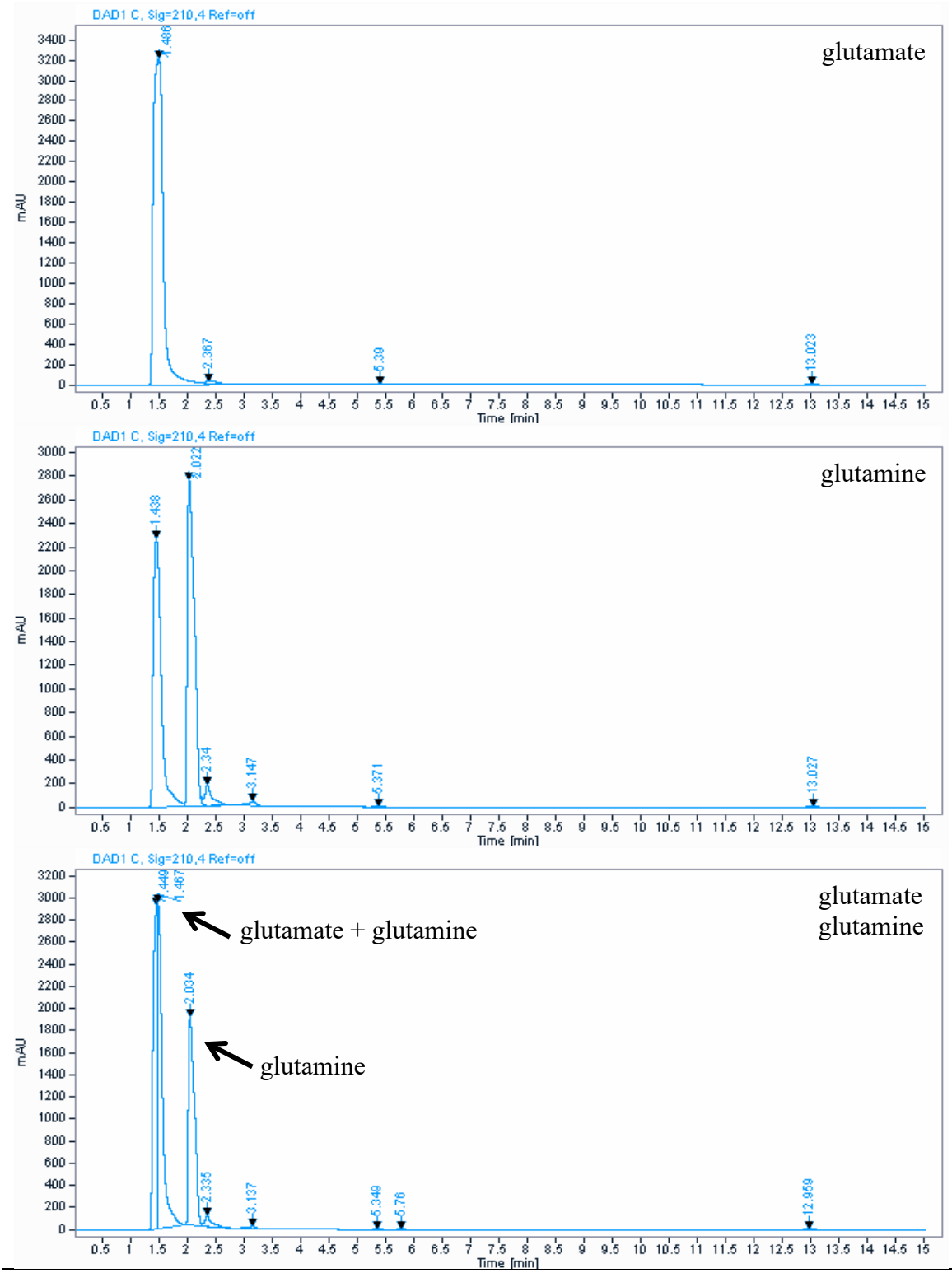


Figure 28: HPLC chromatographs for derivatised glutamate and glutamine at 210 nm show distinct peaks, which can however not be used to discriminate between the two amino acids in mixture due to peak overlap. All samples at 0.5 g/L concentration. n =3

Derivatised sample detected at 338 nm

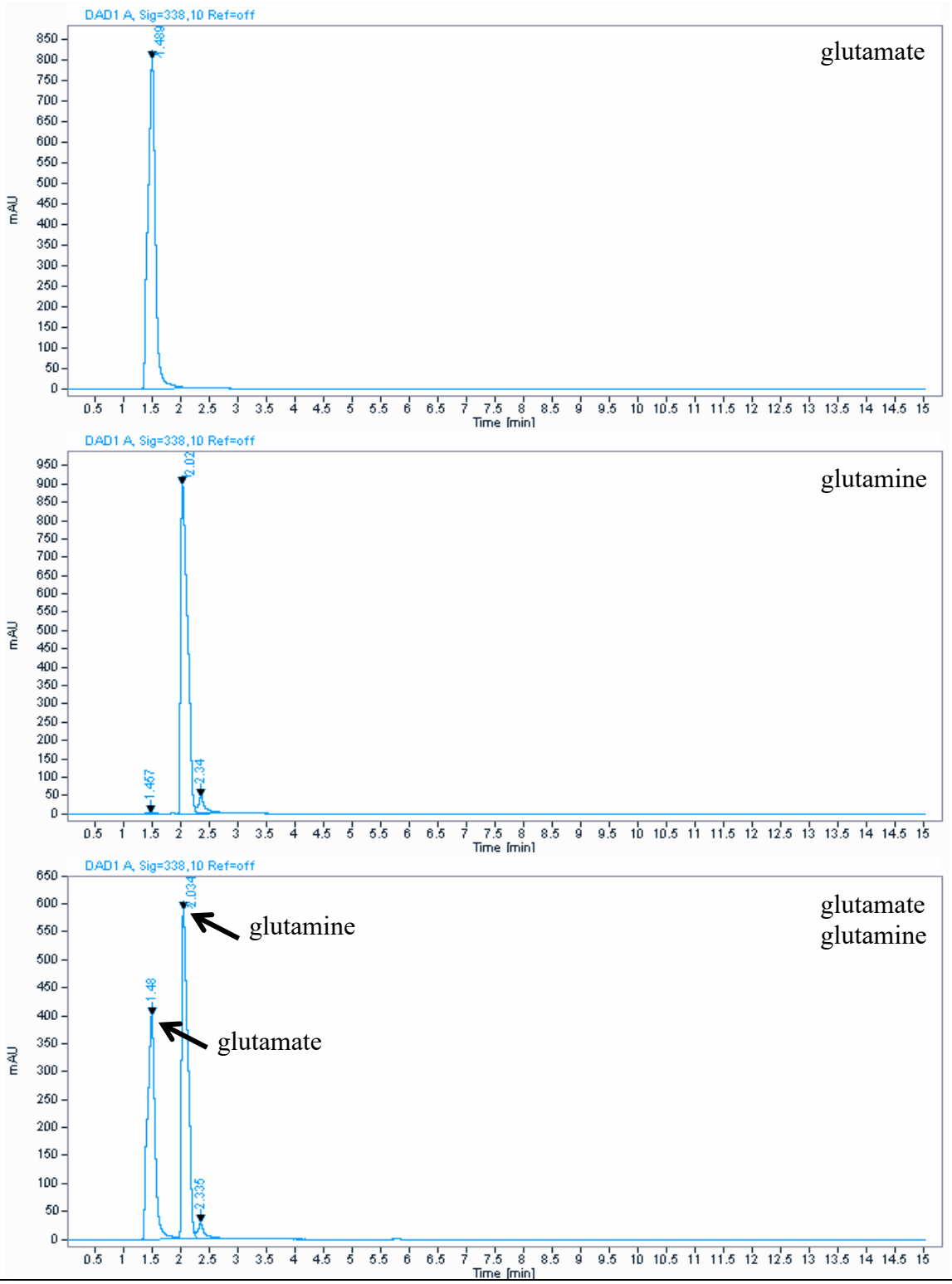
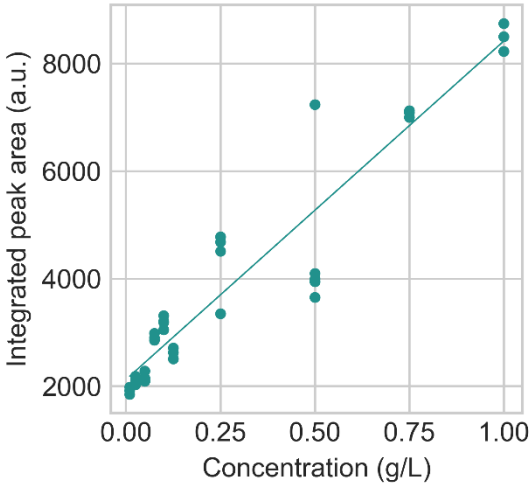


Figure 29: HPLC chromatographs for derivatised glutamate and glutamine at 338 nm show distinct peaks that can be used to distinguish between the amino acids in mixture and quantify them separately. All samples at 0.5 g/L concentration. n =3

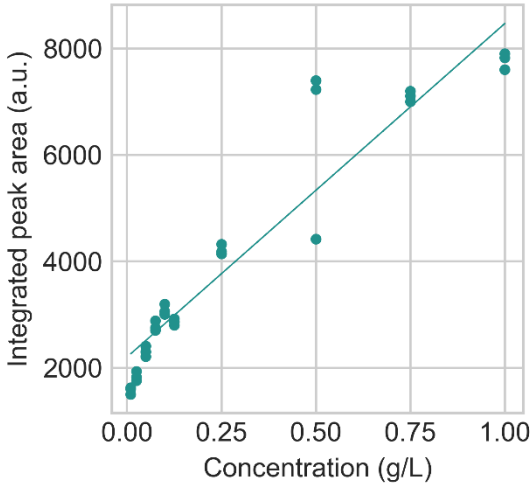
A calibration curve corresponding to concentration of glutamate and glutamine over a range of concentrations between 0 g/L and 1 g/L was prepared to analyse the concentrations of the amino acids in media during coculture experiments (Figure 30).

A) glutamate



$$A = 6295.14 \times c + 2124.17$$

B) glutamine



$$A = 6270.94 \times c + 2196.89$$

Figure 30: HPLC calibration curve with OPA derivatisation detected at 338 nm for A) glutamate ($R^2 = 0.88$) and B) glutamine ($R^2 = 0.88$), where A is integrated peak area (arbitrary units), and c is concentration (g/L).

4.5 Summary

In this chapter, I have undertaken steps to engineer a nutrient exchange consortium mimicking the exchange of nutrients naturally occurring as part of the carbon-nitrogen exchange system within the filament of the multicellular diazotrophic cyanobacterium *Anabaena* 7120 to establish a proof of principle for an obligate symbiosis in which a diazotroph and photoautotroph can provide each other's respective carbon and nitrogen demands. The facultative aerobe *A. vinelandii* was chosen as the nitrogen-fixing partner, and the cyanobacterium *Synechocystis* 6803 as the photosynthetic partner within the consortium.

I have also developed a HPLC protocol for the quantification of glutamate and glutamine, the core amino acids involved in the nutrient exchange pathway investigated in this chapter. Future work could build upon this by validating the HPLC protocol using coculture culture media and establishing the limit of detection (LOD) of the method presented. If necessary, improved peak separation and LOD could be achieved through the use of an alternative HPLC column in follow-up work; ZORBAX Eclipse Amino Acid Analysis HPLC Column (Agilent) for instance is quality tested specifically for the detection and quantification of amino acids when coupled with OPA and FMOC derivatisation.

The work carried out here provides a foundation to setting up and evaluating a nutrient exchange consortium between a photoautotroph and diazotroph. Unfortunately, portions of this experimental section could not be completed as I had to halt all wet lab research due to a sudden onset of illness preventing safe return to lab work.

4.5.1 Transformation of *A. vinelandii*

Genetic parts required for the setup of a GS/GOGAT nutrient exchange system (Figure 19) were designed, synthesised and cloned. Genetic cassettes for the deletion of the nitrogenase

activity regulator (*nifL*) and glutamate synthase (*GOGAT*) were assembled. *A. vinelandii* has been successfully transformed using plasmids carrying *nifL* and *GOGAT* deletion cassettes in a plasmid backbone. However, while growth on selective media was confirmed, it remained unclear whether genetic integration had been achieved. To address this, further experiments can be performed to confirm whether mCherry and second antibiotic resistance was expressed from plasmid or genome using a diagnostic colony PCR with primers corresponding to regions up- and downstream of the region in question.

A. vinelandii is polyploid (Dos Santos, 2011), and integration into all genome copies is expected to happen at different rates. To increase the likelihood of a homozygous genotype, colonies should be passaged on increasing antibiotic concentrations (kanamycin and ampicillin, but without spectinomycin from plasmid backbone), to increase selective pressure. Until then, colony PCR is likely to show multiple bands corresponding to genome copies both with and without the construct. These DNA regions should finally be verified through sequencing. Loss of plasmid is expected to occur in the case of mCherry constructs, as the integrative region carries no selective marker that can be used to apply selection pressure; these constructs served purely to simplify visual readout of successful assembly and transformation. Finally, a double deletion mutant, carrying both a *GOGAT* and *nifL* deletion would be constructed to implement the full exchange pathway. Full set of constructs is summarised in Table 36.

Table 36: *A. vinelandii* constructs under construction in the present work

Description	Construct	Selective marker	Backbone
GOGAT deletion	Δ*GOGAT*::*KanR*	Spectinomycin, Kanamycin	pJM02
nifL deletion	Δ*nifL*::*AmpR*	Spectinomycin, Ampicillin	pJM02
GOGAT + *nifL* double deletion	Δ*GOGAT*::*KanR* Δ*nifL*::*AmpR*	Kanamycin, Ampicillin	pJM02
GOGAT deletion	Δ*GOGAT*::mCherry	Spectinomycin, pink colonies	pJM02
nifL deletion	Δ*nifL*::mCherry	Spectinomycin, pink colonies	pJM02
GOGAT deletion	Δ*GOGAT*::*KanR*	Kanamycin	Chromosome
nifL deletion	Δ*nifL*::*AmpR*	Ampicillin	Chromosome
GOGAT + *nifL* double deletion	Δ*GOGAT*::*KanR* Δ*nifL*::*AmpR*	Kanamycin, Ampicillin	Chromosome

4.5.2 Plasmid Assembly and Transformation of *Synechocystis* 6803

Future work can build upon the research carried out in this chapter by continuing in the cloning and assembly for the transformation of *Synechocystis* 6803 with *ybjL* glutamate transporter and *cscB* sucrose-proton symporter. Expression of constructs in *Synechocystis* 6803 can be achieved from plasmid, for example the broad host-range RSF1010 (Huang et al., 2022; Jain & Srivastava, 2013) and replicons derived from it (Taton et al., 2022) have been demonstrated to have good compatibility with the chassis, including in our group (Yunus,

2019). As *Synechocystis* 6803 is becoming a widely used cyanobacterial chassis, a range of compatible promoters, ribosome binding sites (RBS) and terminators, all of which are required for successful gene expression, are available, and a follow-up project investigating optimal operon configuration would be of interest. Lastly, the stability of transformation can also be improved by genomic integration of introduced genes into a neutral site such as slr0168 alongside a selective marker.

4.5.3 Evaluation and Optimisation of the Consortium

To evaluate the performance of an *A. vinelandii – Synechocystis* 6803 nutrient exchange consortium, a follow-up project will need to culture the consortium in media lacking carbon and nitrogen supplementation. The high rate of respiration characteristic for *A. vinelandii* is a likely challenge to the final success of this consortium. *A. vinelandii* must consume a very high quantity of fixed carbon to not only provide enough ATP for the costly process of nitrogen fixation, but also to support sufficient oxygen scavenging enabling the microaerobic environment vital to nitrogenase activity.

It remains to be seen whether the levels of sucrose production by *Synechocystis* 6803 would be sufficient, given that Burke's medium itself normally contains 20 g/L of sucrose (Dos Santos, 2011). If viability in carbon-free media is not observed at first, a titration of sucrose levels in media could help determine whether any cross-feeding is happening, as even a slight decrease in the sucrose supplementation necessary is indicative of carbon flux from *Synechocystis* 6803 to *A. vinelandii*. Pathway optimisation can further be used to improve sucrose titre. Du et al. (2013) have used a combination of metabolic engineering and salt shock to improve the production of sucrose by *Synechocystis* 6803 to about 5–6 mg sucrose $L^{-1}h^{-1}$, but less than 10% of this was released into the media even after optimisation. Rates of sucrose production as high as 36.1 mg L^{-1} h^{-1} were achieved in another cyanobacterial species

(*Synechococcus elongatus*) (Ducat et al., 2012), which shares the majority of its sucrose biosynthetic pathway with *Synechocystis* 6803, suggesting that further improvements in *Synechocystis* 6803 are possible.

Additional improvements to the consortium can be achieved through further pathway optimisation. An interesting approach to this can be adaptive laboratory evolution (ALE), as both partners within an obligate consortium are under selective pressure to improve the efficiency of cross-feeding. This project involved the deletion of glutamate synthase (GOGAT) in *A. vinelandii*, but as *A. vinelandii* also expresses glutamate dehydrogenase (GDH), this deletion does not result in a glutamate auxotroph. I have made the choice not to pursue auxotrophy to improve strain robustness, but in a downstream project, glutamate auxotrophy can be used to further fortify the dependency of *A. vinelandii* on *Synechocystis* 6803 as a useful tool to improve selection pressure in any downstream ALE pursuits. This work could provide additional insights into the gene regulatory environment surrounding the engineered pathway.

Chapter 5
Evaluation of the Nutrient-Exchange Potential of Rationally Designed Pathways

5.1 Introduction

5.1.1 Rationale

This chapter expands on the idea to model nutrient exchange pathways designed in line with existing nitrogen-carbon exchange systems. In the previous chapter the GS/GOGAT exchange system was modelled and evaluated in terms of yield in different nutrient conditions (Section 4.2.1). Prior to evaluating the growth rates achievable by implementing bioretrosynthetically discovered nutrient exchange pathways, a small subset of rationally designed nutrient exchange pathways was tested. To establish a proof of concept, I have surveyed compounds already present within the chassis as potential carbon and nitrogen-carrying compounds. These compounds are already being synthesised by the organism without any metabolic engineering required.

Rational design was used before attempting to model rare and new-to-nature compounds in an unbiased way to see what the potential of a rational approach is. The potential of rationally designed pathways for nutrient exchange between Synechocystis 6803 and A. vinelandii is here evaluated and contrasted with a high-throughput unbiased approach used in Chapter 6.

5.1.2 Aims

- Individually implement exchange of each compound within the consortium model and evaluate its potential as a nutrient-exchange compound

- Implement a set of rationally designed nutrient-exchange pathways based on native nitrogen metabolism to evaluate their potential as nutrient-exchange pathways

5.2 Evaluation of Maximum Growth Rates Achievable with Nutrient Exchange Pathways

5.2.1 Identification of Candidate Exchange Molecules from a Set of Compounds Native to the Consortium

All intercellular metabolites within the *A. vinelandii* GEM model were surveyed as candidate nitrogen exchange molecules within the consortium system using a computational approach. The consortium model was first constrained to reflect a carbon-replete, nitrogen-free medium, followed by the iterative introduction of a transport reaction of one metabolite at a time between the *A. vinelandii* and *Synechocystis* 6803 cytosols. *In vitro*, this would be equivalent to introducing transporters for the molecule of choice. A positive optimal solution to the *Synechocystis* 6803 biomass objective function – a proxy for growth – was used as indication of successful feeding on the candidate nitrogen source. An analogous implementation in a nitrogen-free, carbon-rich medium was used to identify compounds with carbon-carrying potential from *Synechocystis* 6803 to *A. vinelandii*.

This approach identified 99 candidate carbon carrying compounds and 29 candidate nitrogen carrying compounds from the set of compounds already present in the model. Finally, the sets of *A. vinelandii* compounds capable of supporting *Synechocystis* 6803 growth in nitrogen-unsupplemented media, and the set of *Synechocystis* 6803 compounds capable of supporting *A. vinelandii* growth in carbon-unsupplemented media were combined combinatorically (each carbon-carrying pathway with each nitrogen-carrying pathway) and implemented in a carbon-free, nitrogen-free media simulation. Each pathway combination was subsequently individually optimised for the *Synechocystis* 6803 (Figure 31) and *A. vinelandii* (Figure 32) biomass objective function to estimate possible growth rates.

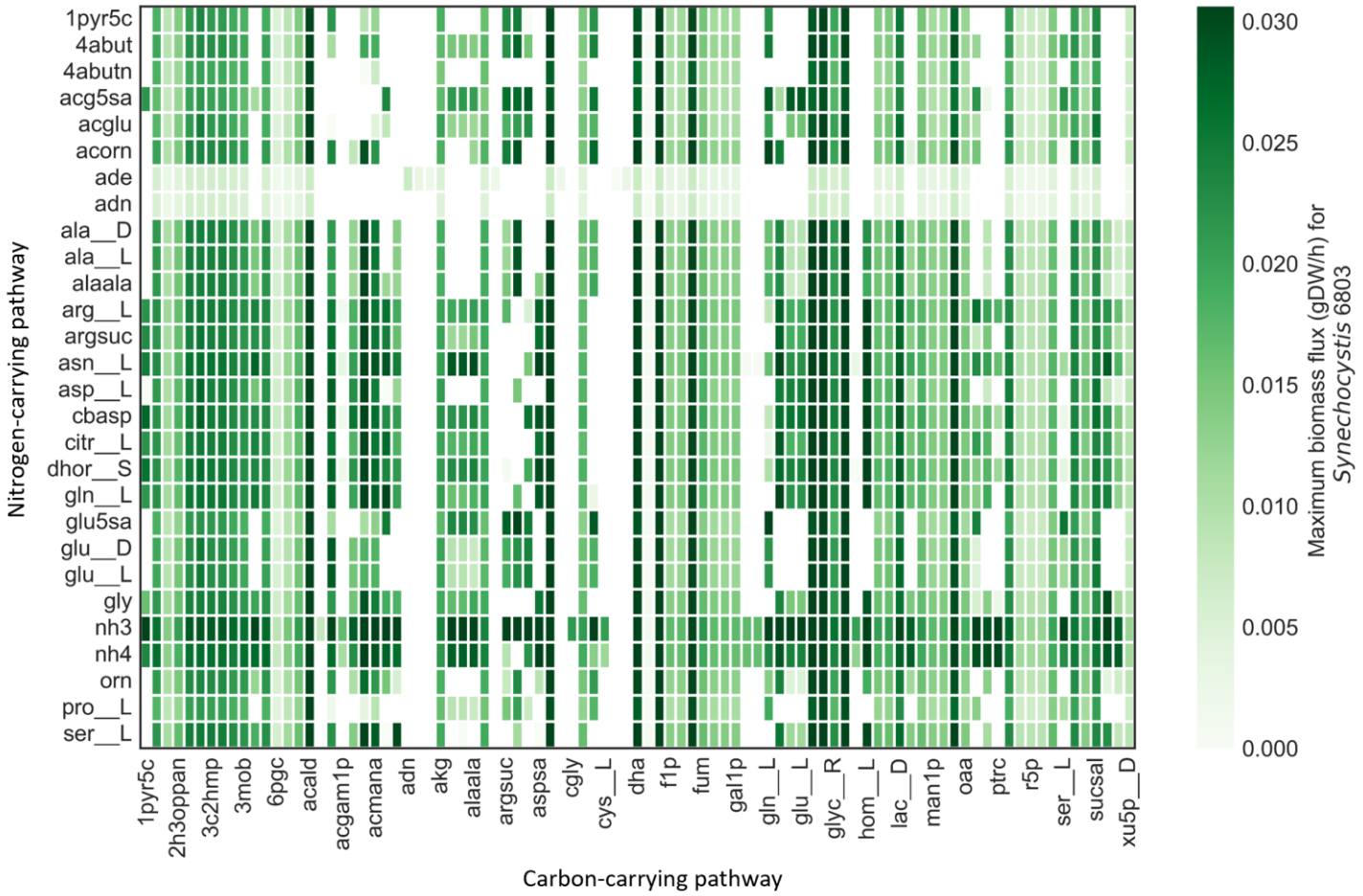


Figure 31: Maximum predicted growth rate of *Synechocystis* 6803 in consortium in nitrogen-free, carbon-free media for each combination of carbon-donor and nitrogen-donor pathways predicted as maximum biomass objective function flux (gDW/h). Compounds are labelled using BiGG identifier for brevity (for compound names, see Supplementary File S3).

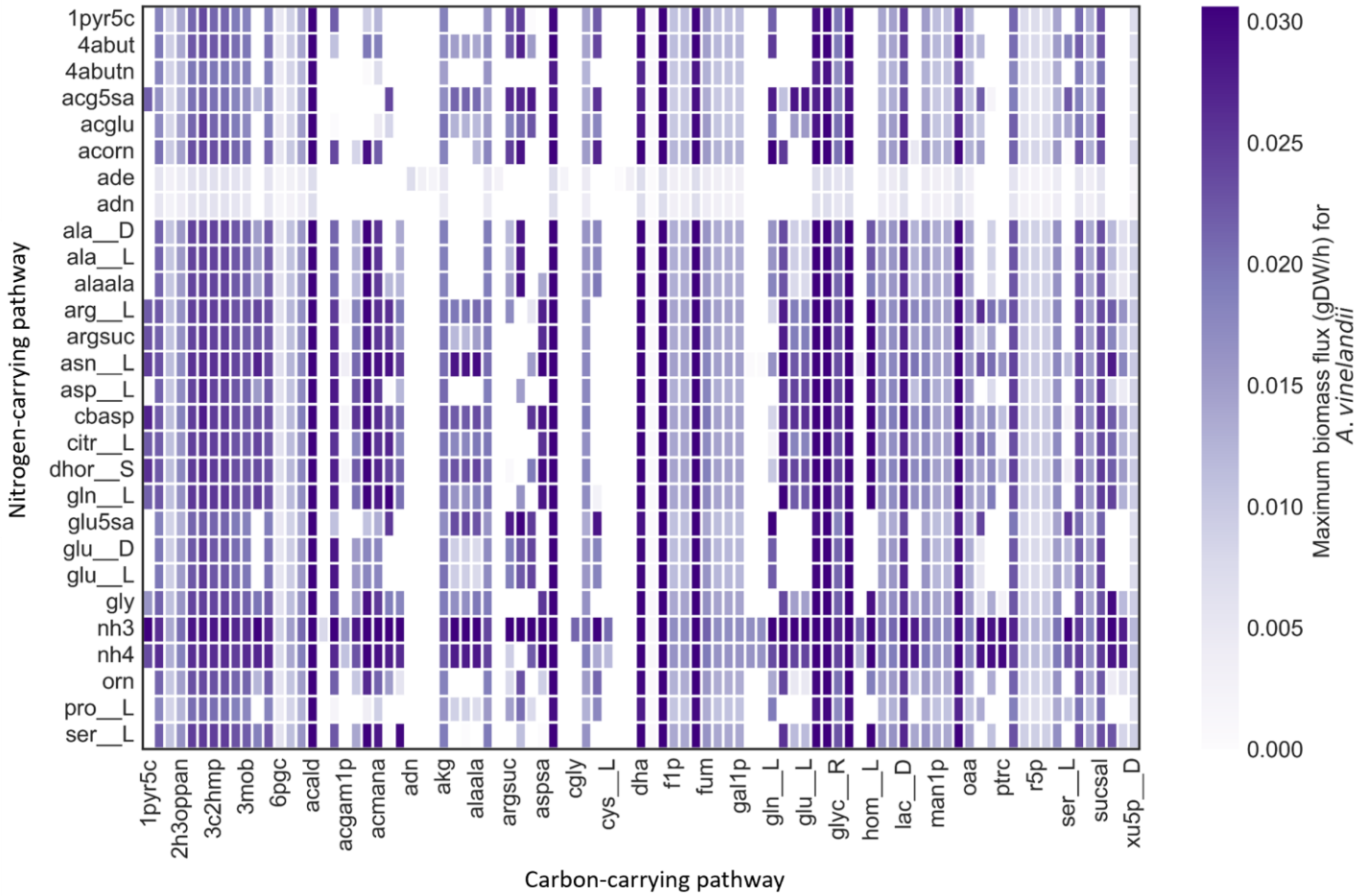


Figure 32: Maximum predicted growth rate of *A. vinelandii* in consortium in nitrogen-free, carbon-free media for each combination of carbon-donor and nitrogen-donor pathways predicted as maximum biomass objective function flux (gDW/h). Compounds are labelled using BiGG identifier for brevity (for compound names, see Supplementary File S3)

5.3 Growth Rates of Consortia with Rationally Designed Nutrient-Exchange Pathways

5.3.1 Deletion of the *A. vinelandii* ammonia reuptake transporter AmtB

Wild-type *A. vinelandii* expresses the membrane-bound ammonia transporter AmtB. AmtB enables the diazotroph to counteract the spontaneous leeching of ammonia into the environment through passive transport; ammonia reuptake plays an essential role in conserving resources invested into nitrogen fixation. An *amtB* deletion has been previously demonstrated to result in increased nitrogen release by *A. vinelandii* (Barney et al., 2015). The *amtB* gene, and the associated *A. vinelandii* ammonia uptake reactions were deleted from the consortium model to simulate this deletion, while adding an ammonia bidirectional transport reaction between the cytosols of *A. vinelandii* and *Synechocystis* 6803 (Figure 33). In this model, ammonia secreted by *A. vinelandii ΔamtB* was able to support the growth of *Synechocystis* 6803 in nitrogen-replete media, and improve its growth rate in rich media relative to wild-type, indicating that nitrogen is limiting growth rate in the wild-type rich media condition (Table 37).

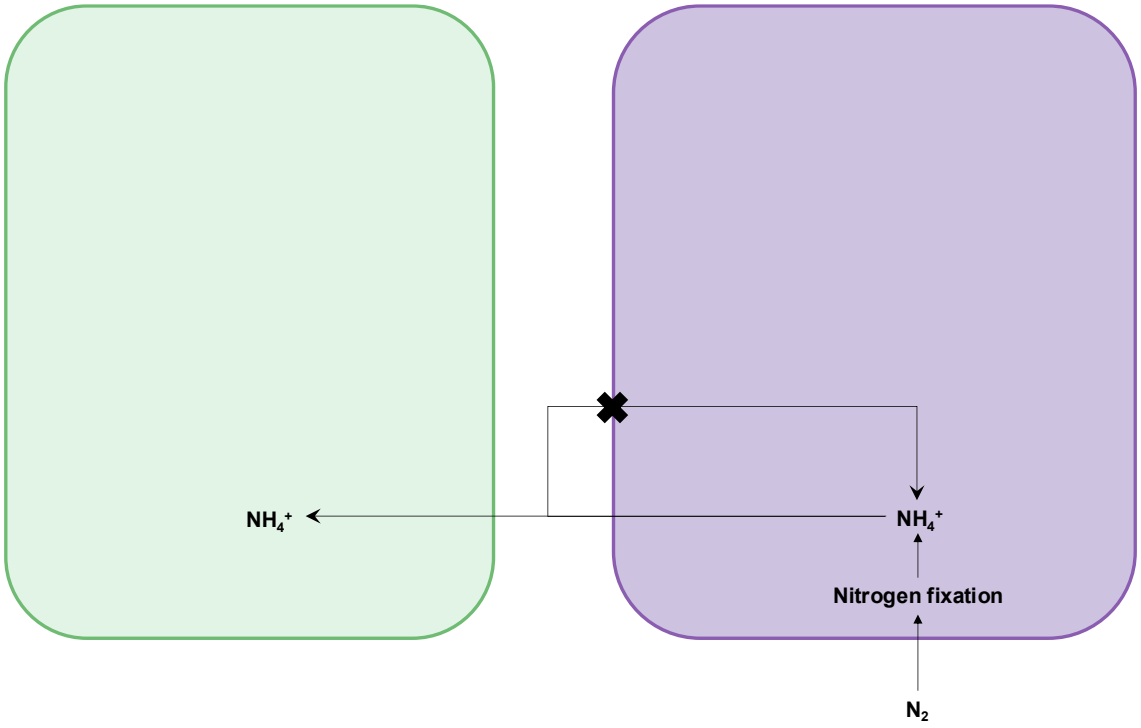


Figure 33: Schematic representation of a nutrient exchange pathway for ammonia involving the deletion of the AmtB ammonia reuptake transporter in *A. vinelandii* and transport of ammonia between cells.

Table 37: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with deletion of *A. vinelandii amtB* per Figure 33.
Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	*ΔamtB*	wild-type	*ΔamtB*
rich	3.45×10^{-3}	3.45×10^{-3}	3.82×10^{-4}	2.17×10^{-2}
no N	3.45×10^{-3}	3.45×10^{-3}	-	2.17×10^{-2}
no C	-	-	3.82×10^{-4}	3.82×10^{-4}
no N, no C	-	-	-	-

5.3.2 *A. vinelandii* Δ*amtB* Combined with a Sucrose-Secreting *Synechocystis* 6803

Deletion of *A. vinelandii* ammonia reuptake transporter was next combined with the introduction of sucrose export from *Synechocystis* 6803. *In vitro*, this corresponds to the introduction of a sucrose exporter such as cscB (Smith & Francis, 2016), as previously employed while implementing a nature-mimicking nutrient exchange in Chapter 4 (Figure 34). Growth rates of both partners are improved relative to the wild-type model in this condition with *A. vinelandii* growth rate improving over two orders of magnitude relative to wild-type. Interestingly, there is a slight improvement in growth rate of *Synechocystis* 6803 in the *ΔamtB* + sucrose relative to *ΔamtB* only, indicating that the photosynthesis flux invested into sucrose supporting growth of *A. vinelandii* not only does not penalise *Synechocystis* 6803 growth rate, but is compensated by the benefit of ammonia received "in exchange", ergo, fixed nitrogen is more limiting to *Synechocystis* 6803 fitness than photosynthesis.

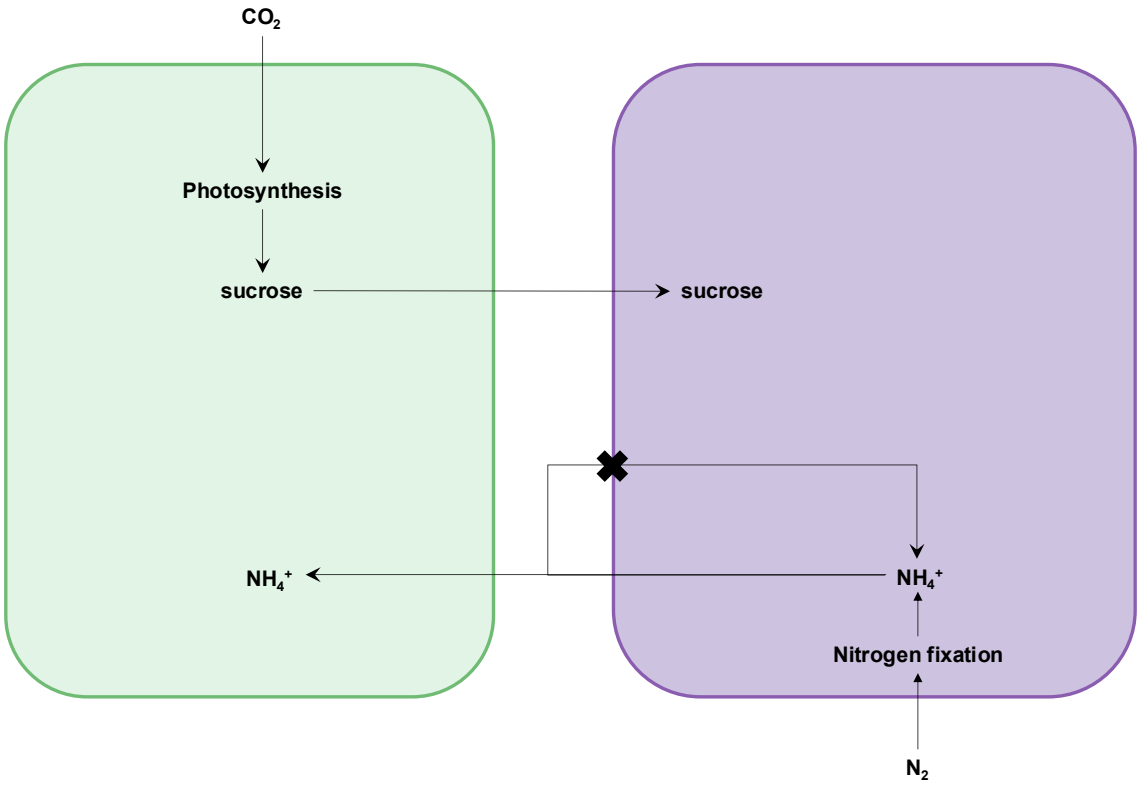


Figure 34: Schematic representation of a nutrient exchange pathway for ammonia involving the deletion of the AmtB ammonia reuptake transporter in *A. vinelandii,* the introduction of a sucrose transporter in *Synechocystis* 6803 and transport of ammonia and sucrose between cells.

Table 38: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with deletion of *A. vinelandii amtB* and export of sucrose from *Synechocystis* 6803 per Figure 34.

Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	*ΔamtB* + sucrose	wild-type	*ΔamtB* + sucrose
rich	3.45×10^{-3}	3.38×10^{-2}	3.82×10^{-4}	2.18×10^{-2}
no N	3.45×10^{-3}	3.38×10^{-2}	-	2.18×10^{-2}
no C	-	2.88×10^{-2}	3.82×10^{-4}	2.17×10^{-2}
no N, no C	-	2.88×10^{-2}	-	2.17×10^{-2}

5.3.3 Exchange of Glutamine and Glutamate

The model was next modified to include exchange reactions for glutamine and glutamate across the two partner cells. This model perturbation does not include any additional modifications of native glutamine and glutamate metabolism (Figure 35). Like in previous predictions, this model supports the growth of both consortium partners in nutrient-limited conditions and improves maximum predicted growth rates even in media that do not lack essential nutrients due to increased availability of limiting nutrients (Table 39). This model was implemented separately with 1) bidirectional exchange of both metabolites – implemented as a reversible exchange reaction with equivalent lower and upper flux constraints[1] – and 2)

[1] Lower flux bound = -1000 mmol/gDWh
Upper flux bound = +1000 mmol/gDWh

directed exchange of metabolites with an irreversible exchange reaction directed from the nitrogen-donating partner *A. vinelandii* for glutamine and from *Synechocystis* 6803 for glutamate. Equal maximum biomass flux was predicted in both cases. This is consistent with the expectation that glutamine, the more aminated of the two amino acids, acts as a nitrogen-carrying molecule, and glutamate, the less aminated of the two, acts as the carbon-carrier; an FBA solution in which this directionality is not the case does not translate to the optimum solution regardless of experimenter-imposed constraints.

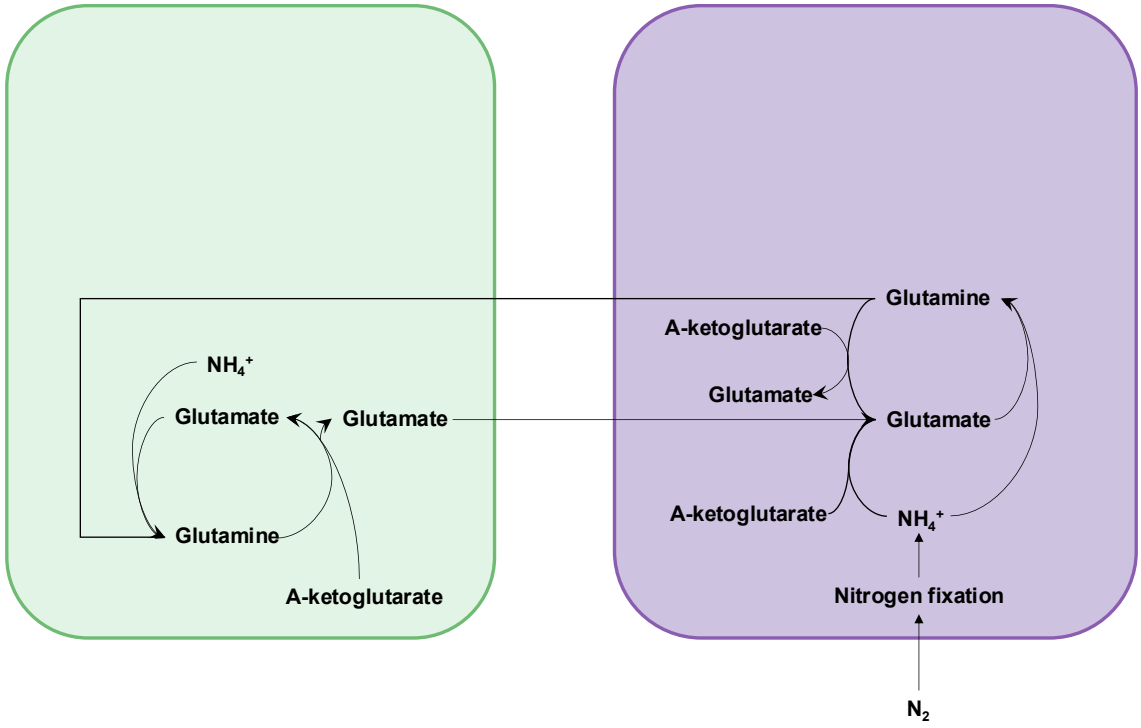


Figure 35: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine and glutamate between *A. vinelandii* and *Synechocystis* 6803.

Table 39: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine and glutamate between cells per Figure 35.

Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	glutamine + glutamate	wild-type	glutamine + glutamate
rich	3.45×10^{-3}	3.34×10^{-2}	3.82×10^{-4}	2.44×10^{-2}
no N	3.45×10^{-3}	3.34×10^{-2}	-	2.44×10^{-2}
no C	-	3.08×10^{-2}	3.82×10^{-4}	2.19×10^{-2}
no N, no C	-	3.08×10^{-2}	-	2.19×10^{-2}

5.3.4 Exchange of Glutamine, Glutamate and Sucrose

The glutamine + glutamate exchange model was next supplemented with a bidirectional exchange reaction for sucrose (Figure 36). Both partners achieve positive growth rates in all media conditions. Consistent with the differences observed between *ΔamtB* (Table 37) and *ΔamtB* + sucrose (Table 38), the growth rate of both partners improves when sucrose exchange is introduced. Similar to the glutamate + glutamine exchange model reported previously, no difference was found between implementations in which the flow of nutrients between cells was directed through the use of flux bounds and undirected.

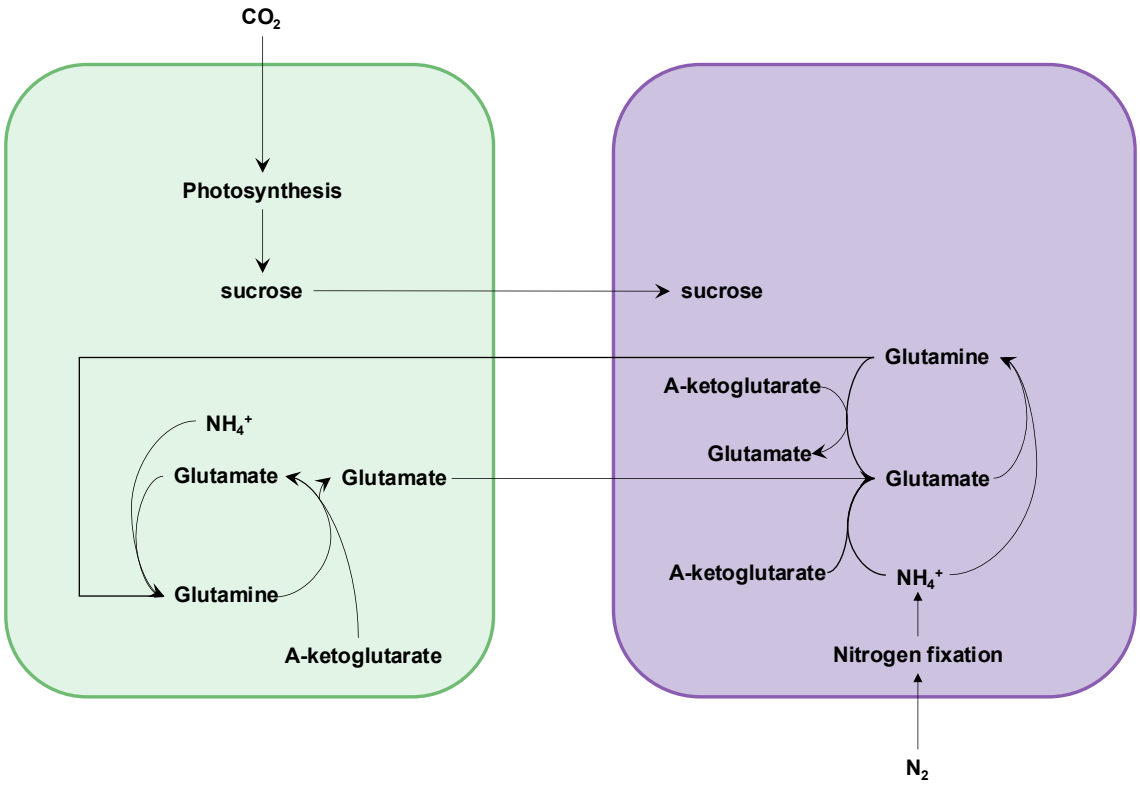


Figure 36: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine, glutamate and sucrose between *A. vinelandii* and *Synechocystis* 6803.

Table 40: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine, glutamate and sucrose between cells per Figure 36.

Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	glutamine + glutamate + sucrose	wild-type	glutamine + glutamate + sucrose
rich	3.45×10^{-3}	3.59×10^{-2}	3.82×10^{-4}	2.53×10^{-2}
no N	3.45×10^{-3}	3.59×10^{-2}	-	2.53×10^{-2}
no C	-	3.08×10^{-2}	3.82×10^{-4}	2.19×10^{-2}
no N, no C	-	3.08×10^{-2}	-	2.19×10^{-2}

5.3.5 Glutamine, Glutamate, Sucrose and A-ketoglutarate

The previously implemented model was further extended to include an exchange of α-ketoglutarate (2-oxoglutarate). In the biosynthesis of glutamate, α-ketoglutarate serves as the carbon backbone to be aminated by GDH and GOGAT enzymes, it thus plays a central role in the assimilation of ammonia (van Heeswijk et al., 2013). The addition of α-ketoglutarate improved the growth rates of both species in all media perturbations (Table 41) relative to glutamine + glutamate + sucrose (Table 40). In all cases, the directionality of α-ketoglutarate exchange flux was observed to be from *Synechocystis* 6803 to *A. vinelandii*, suggesting that α-ketoglutarate synthesised as part of *Synechocystis* 6803 central carbon metabolism can improve

the efficacy of *A. vinelandii* nitrogen assimilation by providing additional backbones for the GOGAT and GDH enzymes.

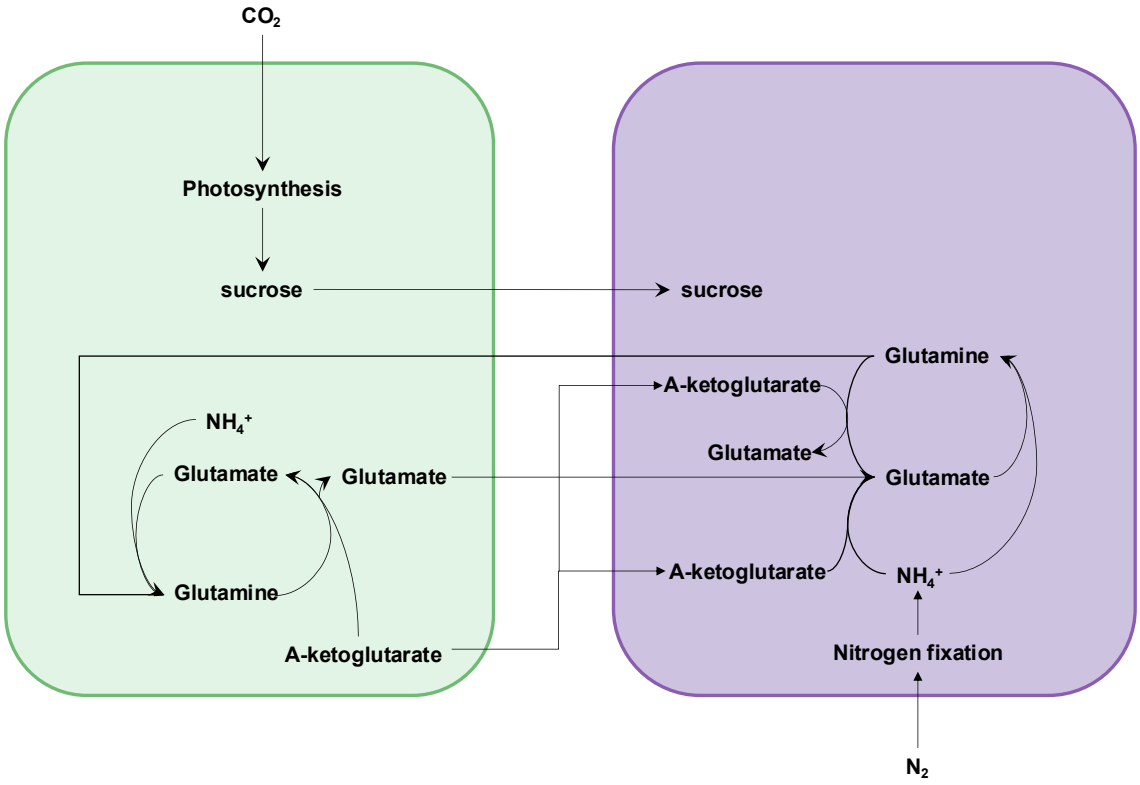


Figure 37: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between *A. vinelandii* and *Synechocystis* 6803.

Table 41: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between cells per Figure 37.

Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	glutamine + glutamate + sucrose + α-ketoglutarate	wild-type	glutamine + glutamate + sucrose + α-ketoglutarate
rich	3.45×10^{-3}	3.88×10^{-2}	3.82×10^{-4}	2.58×10^{-2}
no N	3.45×10^{-3}	3.88×10^{-2}	-	2.58×10^{-2}
no C	-	3.37×10^{-2}	3.82×10^{-4}	2.25×10^{-2}
no N, no C	-	3.37×10^{-2}	-	2.25×10^{-2}

5.3.6 Glutamine, Glutamate, Sucrose and A-ketoglutarate with *ΔgogatΔgdh A. vinelandii*

The glutamine + glutamate + sucrose + α-ketoglutarate model was further expanded by simulating a deletion of the GDH and GOGAT enzymes involved in nitrogen assimilation in *A. vinelandii* (Figure 38). A *ΔgogatΔgdh* phenotype results in a glutamate auxotroph strain of *A. vinelandii*; all glutamate must therefore be acquired from *Synechocystis* 6803.

This model predicts lower growth rates in all media perturbations (Table 42) relative to the strain lacking the deletion (Table 41). As reactions in which α-ketoglutarate is aminated to form glutamate have been removed in this condition, α-ketoglutarate no longer acts as a

nutrient exchange compound, and the predicted flux for α-ketoglutarate exchange is zero in all cases.

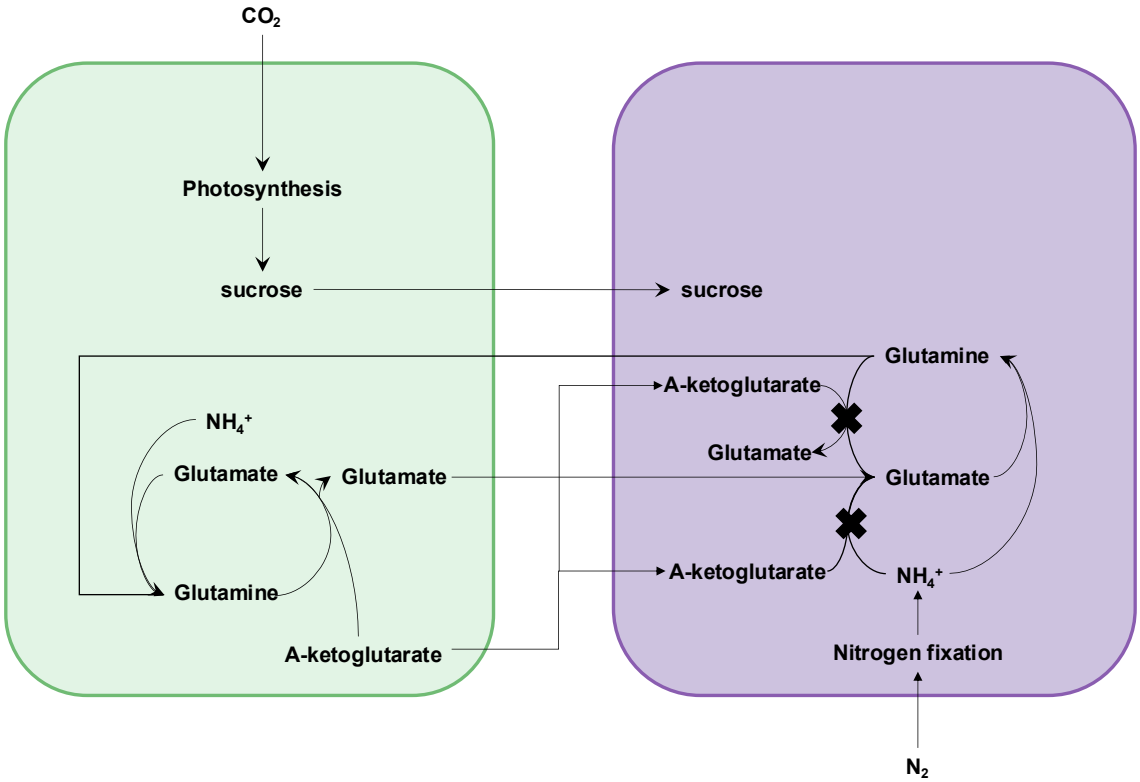


Figure 38: Schematic representation of a nutrient exchange pathway for involving bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between *A. vinelandii* and *Synechocystis* 6803, in addition to a deletion of GDH and GOGAT in *A. vinelandii*.

Table 42: Maximum *in silico* predicted growth rates (gDW/h) of *A. vinelandii* and *Synechocystis* 6803 consortium with bidirectional transport of glutamine, glutamate, sucrose and α-ketoglutarate between cells per Figure 38.

Media included are BNBG11 media perturbations. Zero predicted growth rate indicated by empty cells. Wild-type data originally reported in Table 30 included for comparison.

	A. vinelandii		***Synechocystis* 6803**	
	wild-type	glutamine + glutamate + sucrose + α-ketoglutarate + *ΔgogatΔgdh*	wild-type	glutamine + glutamate + sucrose + α-ketoglutarate + *ΔgogatΔgdh*
rich	3.45×10^{-3}	3.12×10^{-2}	3.82×10^{-4}	2.20×10^{-2}
no N	3.45×10^{-3}	3.12×10^{-2}	-	2.20×10^{-2}
no C	-	2.80×10^{-2}	3.82×10^{-4}	2.19×10^{-2}
no N, no C	-	2.80×10^{-2}	-	2.19×10^{-2}

5.3.7 Summary of Pathways

The maximum growth rates of both *Synechocystis* 6803 and *A. vinelandii* (as predicted using FBA optimising for their respective biomass objective function) in consortium when different nutrient exchange pathways are implemented are summarised in Figure 39, Figure 40.

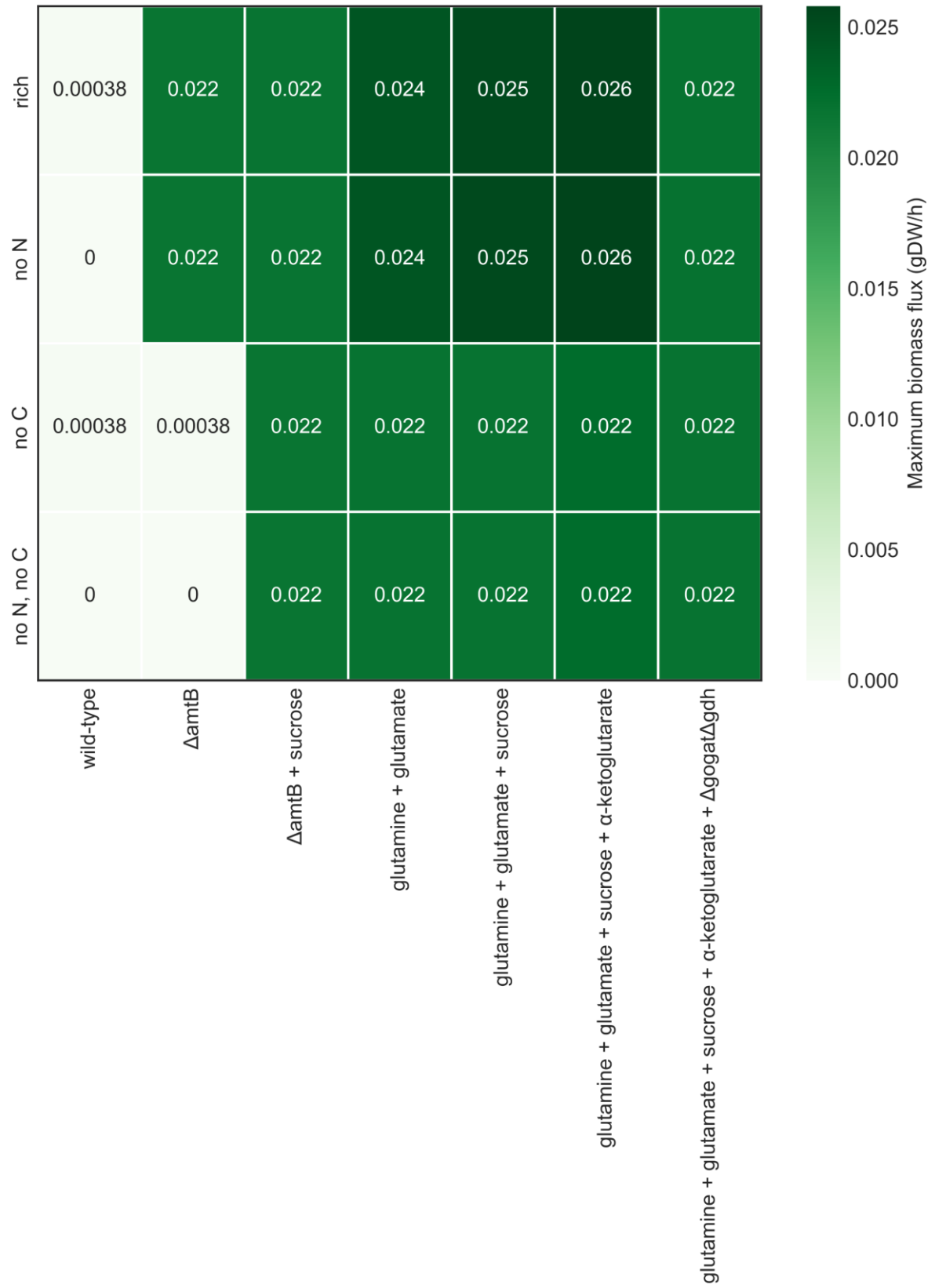


Figure 39: Heatmap of maximum biomass objective function flux for *Synechocystis* 6803 in consortium using different nutrient exchange pathways.

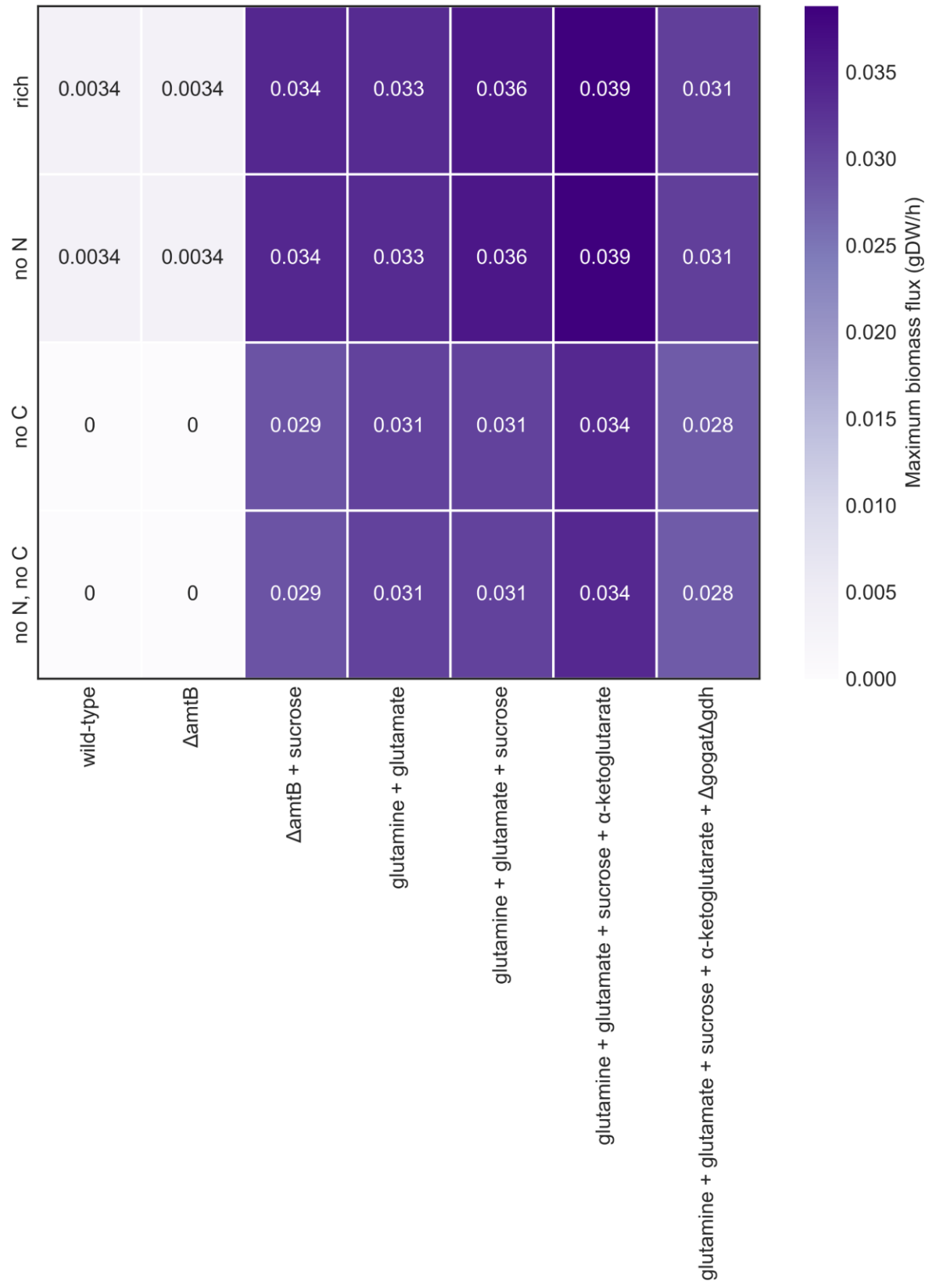


Figure 40: Heatmap of maximum biomass objective function flux for *A. vinelandii* in consortium using different nutrient exchange pathways.

5.4 Summary

In this chapter, I have used FBA to evaluate the potential of using a small number of naturally occurring compounds as nutrient exchange molecules in an *A. vinelandii-Synechocystis* 6803 consortium. These nutrient exchanges were implemented as simple bidirectional exchange reactions without any additional optimisation steps and tested first separately in C-free or N-free media, followed by a combinatorial approach producing a total of 1,901 nutrient exchange pathways with non-zero growth for both partners with *Synechocystis* 6803 growth rates ranging up to 3.60×10^{-2} gDW/h and *A. vinelandii* growth rates up to 5.18×10^{-1} gDW/h in C-free, N-free media. It is to be noted that the exchange of nutrients between *A. vinelandii* and *Synechocystis* 6803 was previously shown to be non-existent or negligible *in vitro* (Figure 15, Figure 16), as the wild-type organisms do not enable each other to grow in media deplete for each partner's respective essential nutrient. Ergo, to implement *in vitro* any exchange that has been determined theoretically compatible with nutrient exchange in this chapter, efficient transport of nutrients across the cell membrane would need be engineered.

Additionally, 6 different nutrient exchange pathways based around central nitrogen metabolism and specifically utilising the GDH and GOGAT enzymes were implemented and tested, procuding maximum predicted growth rates of up to 2.25×10^{-2} gDW/h for *Synechocystis* 6803 and 3.37×10^{-2} gDW/h for *A. vinelandii*. The findings reported here suggest that a nutrient-exchange symbiosis is feasible in an *A. vinelandii-Synechocystis* 6803 consortium, and that both partners can theoretically achieve higher growth rates when cultured together than alone, suggesting a mutually beneficial mutualistic symbiosis.

Chapter 6
Design of Nutrient Exchange Pathways with Bioretrosynthesis

6.1 Introduction

6.1.1 Rationale

In a nutrient-exchange consortium, partners rely on each other to synthesise and exchange compounds containing nutrients essential to the survival of the other partner in a given environment. In this study, I am looking to identify compounds carrying bioavailable nitrogen that can be synthesised by the diazotrophic bacterium *A. vinelandii*, and assimilated by the photoautotrophic cyanobacterium *Synechocystis* 6803, and carbon-containing organic compounds synthesisable by *Synechocystis* 6803 and, in exchange, used by *A. vinelandii* so that a consortium of the two partners achieves survivability in media lacking supplemented fixed carbon or nitrogen.

In this chapter, candidate exchange compounds are evaluated *in silico*. A computational approach enables a high throughput, where thousands of compounds and pathways can be tested at a significantly reduced time and cost investment than *in vitro*. A wide scope also means that more exotic compounds will be taken into consideration. Earlier in this thesis, I have proposed that inefficiencies in the exchange of nutrients between plants and soil-dwelling free-living nitrogen-fixers can, in part, be explained by the uptake of released nutrients by other members of the ecosystems. We hypothesise that compounds that do not commonly occur in natural ecosystems may be attractive synthetic exchange compounds precisely because they are less likely to be available as nutrients to other members of the rhizosphere.

The first criterion a candidate nutrient exchange compound needs to satisfy is for it to be 1) possible to synthesise in the donor chassis and 2) possible to break down by the recipient

chassis. A pathway prediction approach was chosen to evaluate nitrogen or carbon-carrying compounds for which such pathways could be engineered. To achieve this, I used a published pathway prediction algorithm to design candidate biosynthesis and breakdown pathways for a large set of a carbon- and nitrogen-containing compounds retrieved from a biochemistry database.

6.1.2 Aims

- Identify a set of candidate exchange compounds
- Design biosynthesis and breakdown pathways for a set of candidate exchange compounds
- Rank proposed pathways based on their feasibility

6.2 Selection of Candidate Compounds

Candidate exchange compounds were chosen from the total set of unique compounds contained within the MetaNetX database (https://www.metanetx.org/) (Bernard et al., 2014; Ganter et al., 2013; Moretti et al., 2016a; Moretti et al., 2021). The MetaNetX database contains just under 1.3 million compounds included in a range of genome-scale metabolic model reconstructions. Entries are curated to link compound representations within metabolic models with their respective chemical identity and structural information, which enables standardisation across models, reconciling metabolites which may be denoted using alternate representations in various publications, but in fact represent the same chemical entity.

For the purposes of this project, this database was chosen as it contains many diverse compounds, which are still in close proximity to the natural biological scope to have previously been included in some published genome scale metabolic model. The majority of the included compounds are also organic compounds, which therefore more easily plug into the catalytic capabilities of living organisms. Finally, this database was well suited to programmatic conversion between unique chemical identifiers and compound aliases used in genome-scale metabolic models used later in this project (Section 5.3.7).

All compounds within the MetaNetEx database were retrieved. Entries lacking a chemical formula and InChI identifier were excluded (Table 43). The InChI is a standardised unique alphanumeric representation of a chemical formula with bond connectivity which includes stereochemistry, tautomerism, isotope and charge information (InChI Trust, 2024). This was then used to filter for unique compounds. The compounds were further subset into those containing carbon and those containing nitrogen in their formulae; the final sets overlap on compounds containing both nutrients. Full list of carbon and nitrogen-containing compounds queried is available in Supplementary files S1 and S2, respectively.

Table 43: Subsets of compounds retrieved from MetaNetX database used in this analysis

Description of Set	**Number of Compounds**
Total Number of Compounds in MetaNetX	1,292,154
Unique Compounds with Chemical Formulae and InChI	1,049,424
Carbon-containing candidate compounds	1,048,228
Nitrogen-containing candidate compounds	381,616

6.3 Bioretrosynthesis-Based Pathway Prediction

A bioretrosynthesis approach was chosen to design biosynthesis and breakdown pathways for each candidate nutrient exchange compound. Retrosynthesis is a framework within synthetic chemistry to design *de novo* compound synthesis routes. A set of feasible chemical transformations is applied to the compound of interest, yielding a set of potential intermediates, used in the next iteration of chemical transformations. This process is repeated until a commercially available and easy-to-source compound is found. The route identified is then reversed to yield a synthetic process.

In the context of metabolic engineering, bioretrosynthesis seeks to identify the enzymatic steps required to build a path from the desired metabolite back to a set of compounds available in the background chassis. One major advantage of retrosynthesis over forward synthesis is a decrease in computational complexity, where a smaller number of molecules needs to be evaluated at each step when starting from the target, rather than starting from all the possible available precursors.

6.3.1 RetroPathRL: Background on the Bioretrosynthesis Tool Used

RetroPathRL (Koch et al., 2019) is an open-source tool developed for computational bioretrosynthesis building on previously released RetroPath2.0 (Delépine et al., 2018) and RetroPath (Carbonell et al., 2013). The RetroPathRL software applies a set of reaction rules compiled from known enzymatic and non-enzymatic chemical transformations within biological systems, derived from a range of biochemical databases (Duigou et al., 2019). Reaction rules depict the change in bonding patterns between product and substrate. A rule-based approach to retrosynthesis explores a wider solution space than if only known

biochemical transformations were included, which enables it to account for enzyme promiscuity. Enzyme promiscuity is known to play a significant role in biochemical processes, in *E. coli*, for example, 37% of enzymes have been estimated to have promiscuous activity (Nam et al., 2012). Protein engineering is also increasingly combined with metabolic engineering approaches; this can be further used to alter or shift the substrate specificities of enzymes within a pathway. In this context, pathways predicted with a rule-based approach can contain compounds outside of biochemical databases, and potentially novel to nature.

6.3.1.1 Reaction Scoring

Reactions are scored at every step of tree expansion based on two criteria: chemical and biological score. First, for the chemical score, Tanimoto similarity is calculated between query and native substrate first before rule application, and later between query and native products after rule application. The term Jaccard similarity is also used to refer to Tanimoto similarity. Tanimoto similarity is a commonly used measure of chemical similarity representing the ratio of the intersection of two bit vectors over their union (Bajusz et al., 2015; Willet et al., 1998):

$$T(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

In the case of cheminformatics, these bit vectors are molecular fingerprints encoding the 2D conformation of atoms in a molecule in the form of an integer array. RetroPathRL specifically uses Morgan fingerprints implemented as the GetMorganFingerprintAsBitVector() function in the RDKit Python package (Laudrum & others, 2024; Rogers & Hahn, 2010). In other words, the structural formula of both the reactants and the products in a predicted reaction are compared to the reactants and products of the naturally occurring reaction the applicable rule derives from. These similarities are then used to derive the chemical score for a given reaction on a scale from 0 to 1, with 1 representing the highest confidence.

Next, the biological score represents the likelihood that an enzyme catalysing a reaction which satisfies the reaction rule for a step can be found. This scoring system was devised by Delépine et al. (2018). The authors retrieved enzyme sequences from the UniProt database (The UniProt, 2019) and compared these with annotated reactions in the Rhea (Morgat et al., 2017), MetaCyc (Caspi et al., 2018) and Reactome (Fabregat et al., 2018) databases to assign each sequence to the applicable reaction rule, or the type of reaction catalysed. Both sequences and reaction rules were independently clustered by similarity. The penalty score was calculated as:

$$penalty_{score}(rule) = \log_{10}(n_{rule})$$

where n_{rule} is the number of distinct clusters containing sequences associated with the rule. A lower penalty score means fewer sequence clusters are associated with a reaction rule; the authors interpret this as translating to a higher confidence in sequence-rule relationship. The penalty score is then converted and normalised to yield a biological score on a scale from 0 to 1, with 1 representing the highest confidence.

Finally, RetroPathRL uses a biochemical score, produced by multiplication of biological and chemical scores, to determine reaction "trustworthiness" (Koch et al., 2019):

$$score_{biochemistry} = score_{biology} \times score_{chemistry}$$

6.3.1.2 Tree Search

RetroPathRL uses the Monte Carto Tree Search (MCTS) reinforcement learning algorithm. MCTS has previously been used by Segler et al. (2018) to design retrosynthesis pathways in the context of synthetic chemistry, achieving a 30-fold improvement in runtime when compared with traditional retrosynthesis software. This reinforcement learning algorithm explores an unknown space using random sampling in four repeated phases: selection, expansion, simulation and backpropagation (Figure 41). During selection, best child nodes are iteratively selected from the root node until a previously unvisited leaf node is reached.

RetroPathRL uses the biochemical score as selection policy. Next, a leaf is expanded by generating a predefined number of its children based on a set of possible transformations defined by the selected ranking scheme (biochemical score). Simulation is an iterative procedure in which the state of the leaf is first evaluated. If a leaf is terminal – having reached the maximum depth or a compound present in the background chassis – a penalty or reward is applied, respectively. Else, a random feasible transformation is applied, extending the tree, and the simulation step is repeated until the maximum depth or maximum number of rollout steps is reached. Finally, the score obtained after exploring a node is backpropagated, updating values and visit counts of parent nodes (Koch et al., 2019).

One major advantage of the RetroPathRL software is its open-source nature. All code is freely available and fully documented in a dedicated GitHub directory (https://github.com/brsynth/RetroPathRL). A user therefore does not need to implicitly trust the software while it is operating as a black box, enabling improved traceability of experiments. Additionally, any parameter may be tuned by an expert user (Koch et al., 2019). As a python package, the software can also be easily installed locally or on a high-performance cluster to easily parallelise computations, in contrast with many other web-based pathway design applications, or even the previous release of RetroPath, which relied on the KNIME Analytics Platform (Berthold et al., 2008; Delépine et al., 2018).

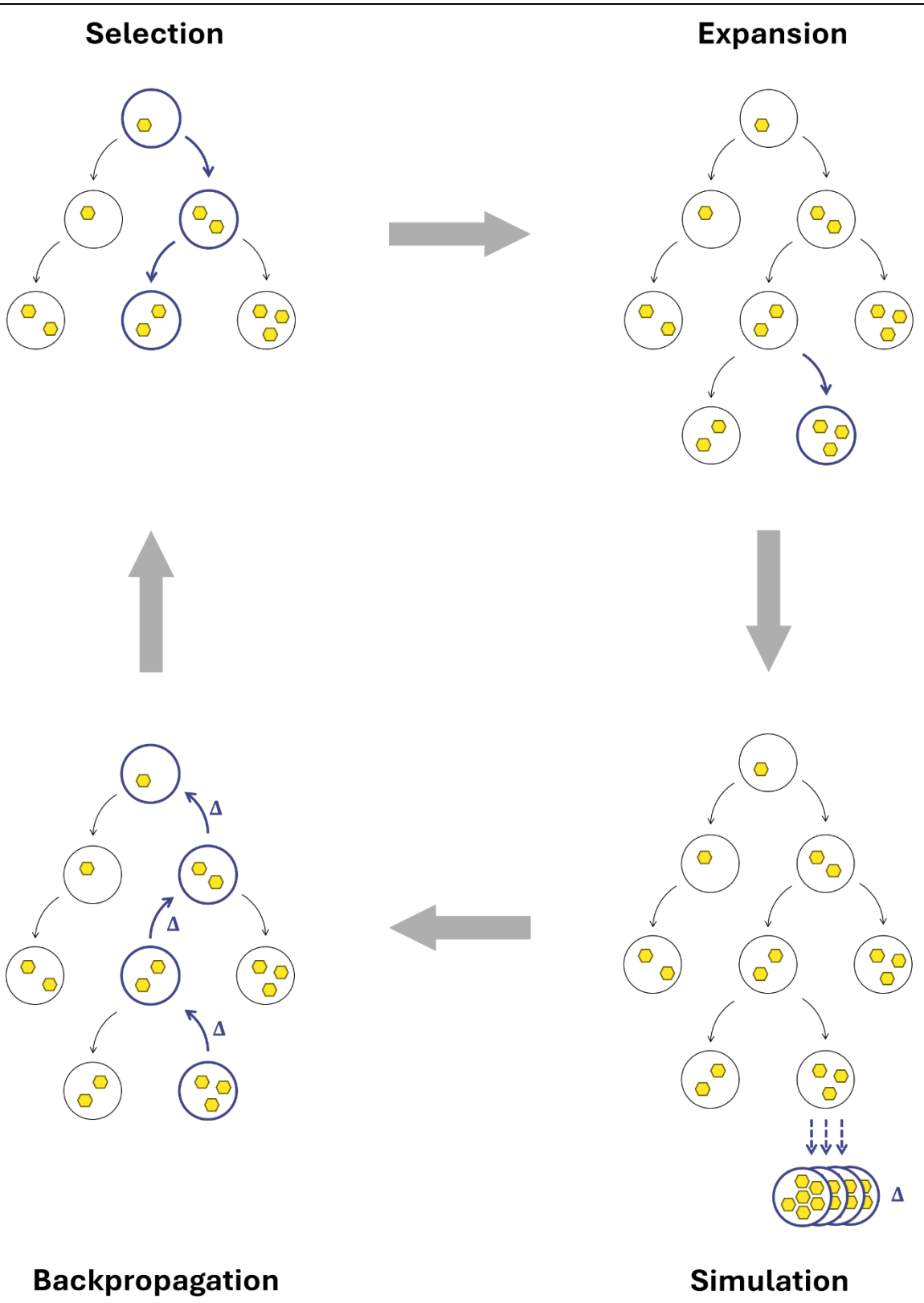


Figure 41: Schematic representation of the phases of the Monte Carlo Tree Search Reinforcement learning method. Adapted from Koch (2019)

6.3.2 Implementation and Parameters

A unique script defining RetroPathRL parameters was generated for each compound of interest for both chasses. Query compounds were specified using their respective InChI identifiers retriever from the MetaNetX database as outlined in section 6.2. The list of compounds available in chassis was derived from the list of compounds in the *A. vinelandii* and *Synechocystis* 6803 models used in this thesis – iAA1300 (Alleman et al., 2021) and iSynCJ816 (Joshi et al., 2017), respectively. Compounds were converted from Bigg identifiers into InChI as outlined in Materials and Methods, Section 2.3.1. Full set of parameters is outlined in Materials and Methods, Section 2.3.4, Table 20. In general, parameters were chosen following best practices outlined by Koch et al. (2019).

The RetroPathRL parameter diameter specifies the degree of promiscuity allowed in a reaction rule, representing the number of atoms at distance one around the reaction centre minus one which are considered in the construction of reaction rules. The reaction centre is the portion of the molecule which changes its configuration when a reaction is applied. I have carried out exploration at diameters 10, 12, 14 and 16 (medium to high specificity). Specificity lower than 10 was not included in this analysis to limit expansion of design space – a lower diameter results in higher branching factor and thus more potential pathways per candidate compound – and to bias results toward lower enzyme promiscuity.

A maximum pathway length of 7 was selected to balance the number of available enzymatic transformations and the feasibility of *in vivo* implementation. More exotic compounds are likely to require several catalytic steps, thus favouring longer maximum pathway length. Pathways consisting of many enzymes are also less likely to spontaneously arise in competing organisms within a natural ecosystem, which improves the isolation of the engineered nutrient exchange system from the environment, decreasing the likelihood of

nutrient leakage. On the other hand, while there is no limit to the length of a pathway *in silico*, real-world metabolic engineering favours shorter pathways. While recent advances in DNA assembly methods continue increasing the number of DNA parts which can be successfully engineered – for instance BASIC assembly, which has been used as part of this project, has been shown to achieve 90% accuracy with 7 part assembly (Storch et al., 2015) –, the introduction of more genetic parts and protein expression is associated with a growth rate penalty and metabolic as well as gene expression burden (Lynch & Marinov, 2015).

This approach was used to identify synthesis as well as breakdown pathways, with the caveat that pathway directionality is later reversed for breakdown pathways. At least one pathway was identified for 7,775 carbon-containing compounds in *A. vinelandii* and 5,488 carbon-containing compounds in *Synechocystis* 6803; for 3,693 compounds, a pathway was found in both species (Figure 42). For nitrogen-containing compounds, at least one pathway was identified for 1,490 in *A. vinelandii* and 1,509 in *Synechocystis* 6803, with an overlap of 1,488 nitrogen-containing compounds for which a pathway was found in both (Figure 43). Each prediction includes unique identifiers of intermediates, enzyme EC numbers, biological and chemical scores, and reaction directionality required to achieve biosynthesis; the results, however, do not account for atomic balance.

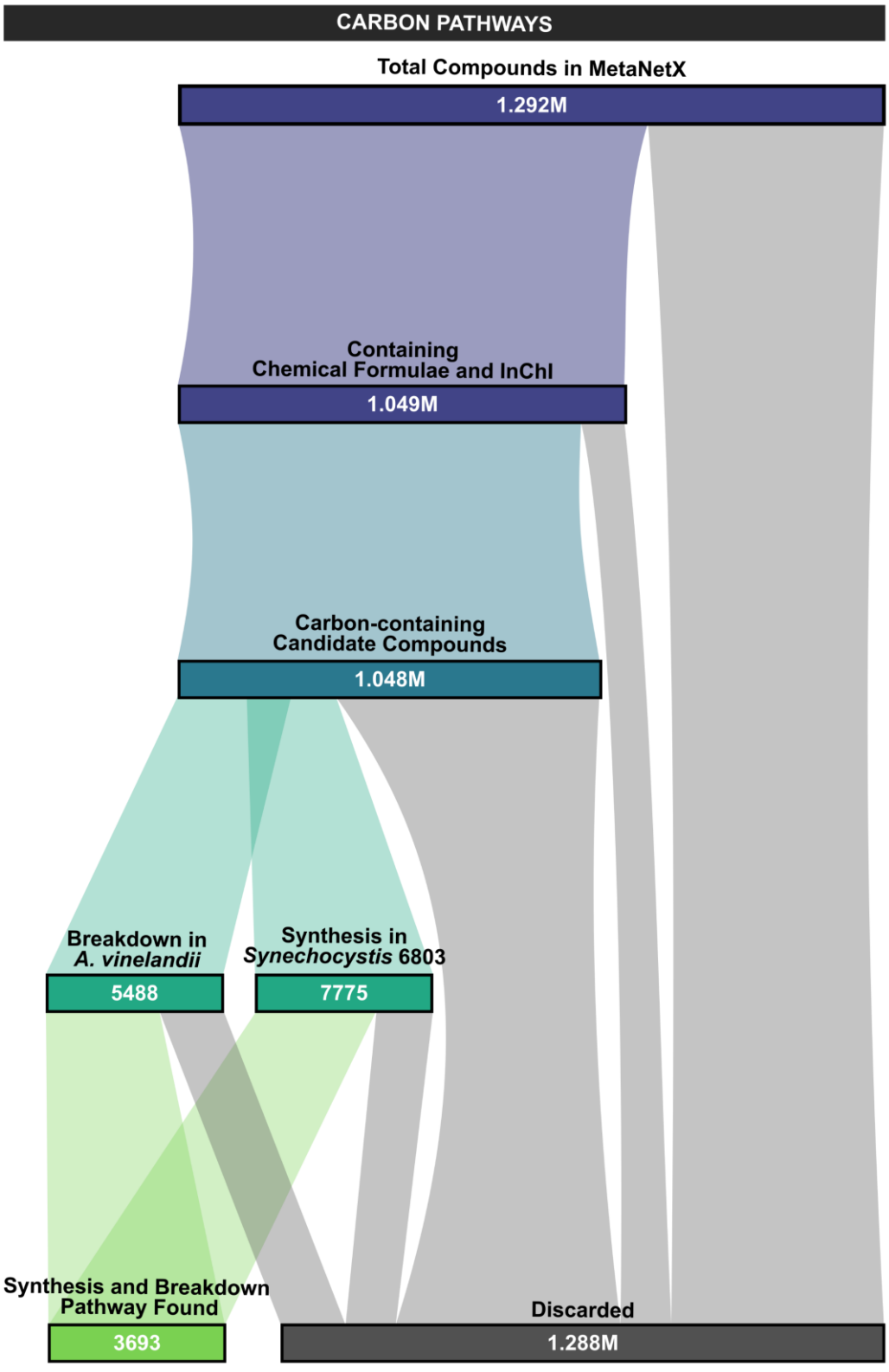


Figure 42: Flow diagram representing the selection process of carbon-containing compounds from the MetaNetX database to those which could be synthesised and broken down in the respective chasses.

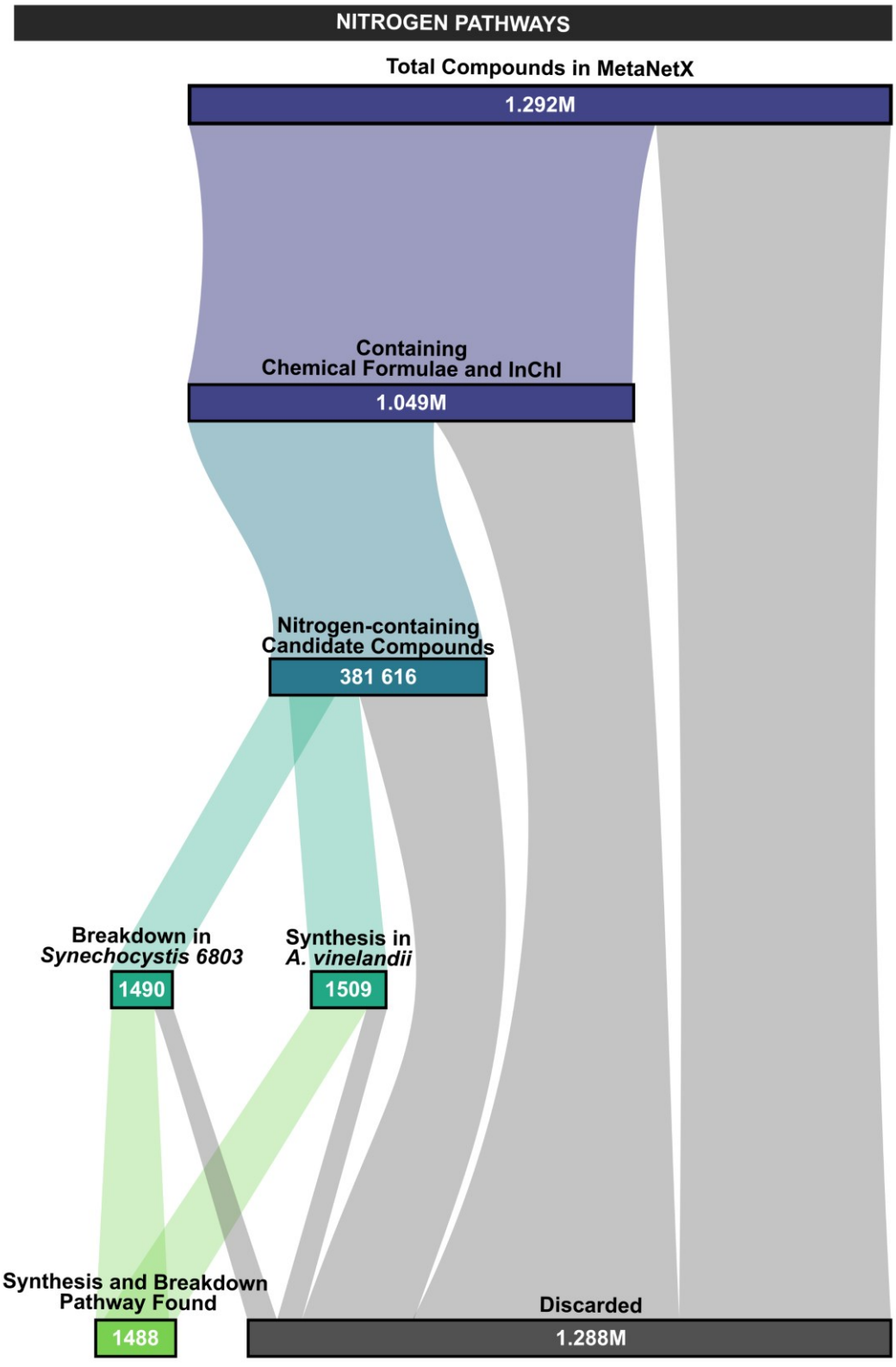


Figure 43: Flow diagram representing the selection process of nitrogen-containing compounds from the MetaNetX database to those which could be synthesised and broken down in the respective chasses.

6.3.3 Pathway Combinations and Scoring

Finally, the predicted pathways for the biosynthesis and breakdown of each compound were combinatorically joined so that for each compound, every predicted biosynthesis pathway was combined with every predicted breakdown pathway. This has yielded 139,139 different pathway matches for nitrogen-exchange compounds and 294,331 pathway matches for carbon-exchange compounds. Predicted pathways were finally ranked based on their biochemical score and pathway length. The pathway biochemical score was calculated by multiplication of biological and chemical scores of each reaction, averaging across the entire pathway.:

$$score_{pathway\ biochemistry} = \frac{1}{n}\sum_{i=1}^{n}(score_{biology} \times score_{chemistry})$$

where n is number of reactions in a pathway.

A priority score was calculated as pathway biochemical score divided by pathway length:

$$score_{priority} = \frac{score_{pathway\ biochemistry}}{n}$$

This priority score was used to determine the order of execution of downstream simulations (Chapter 7) to prioritise simulating the pathways seen as most promising based on lowest pathway length and highest score.

The generated pathways are summarised in Figure 44 (nitrogen-carrying pathways) and Figure 45 (carbon-carrying pathways) ranked based on biochemical score and pathway length. The mode length and score for nitrogen-carrying pathways corresponds to a length of 10 and score 0.6 (Figure 44), and for carbon-carrying pathways the mode is at length of 7 and score of 0.8 (Figure 45). This suggests that a higher proportion of carbon-carrying pathways can be successfully implemented than in the case of nitrogen-carrying pathways, as shorter pathways are easier to engineer, and a higher score represents higher pathway "trustworthiness".

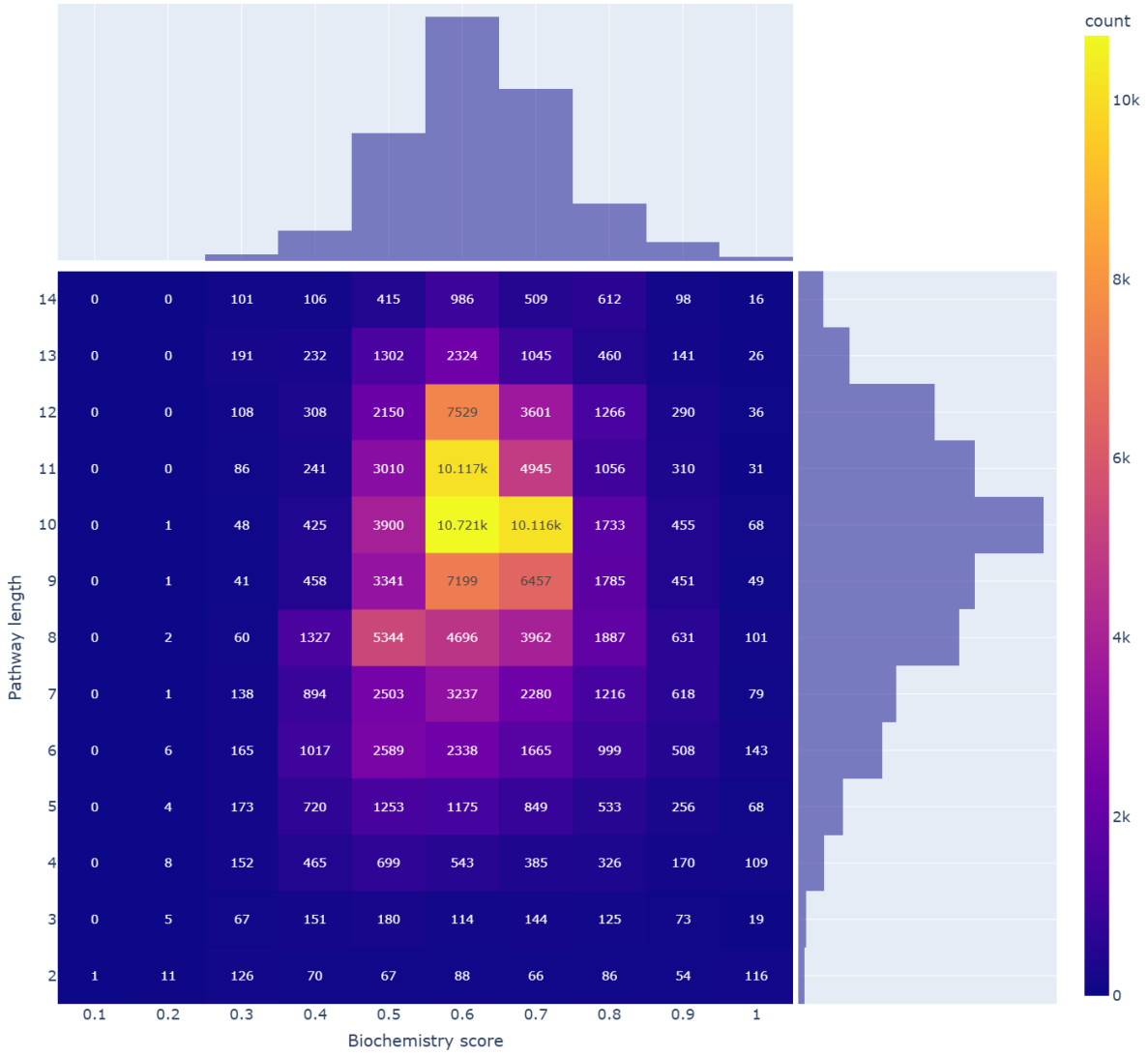


Figure 44: Heatmap representing the pathway lengths and biochemical scores of the designed nitrogen-carrying pathways.

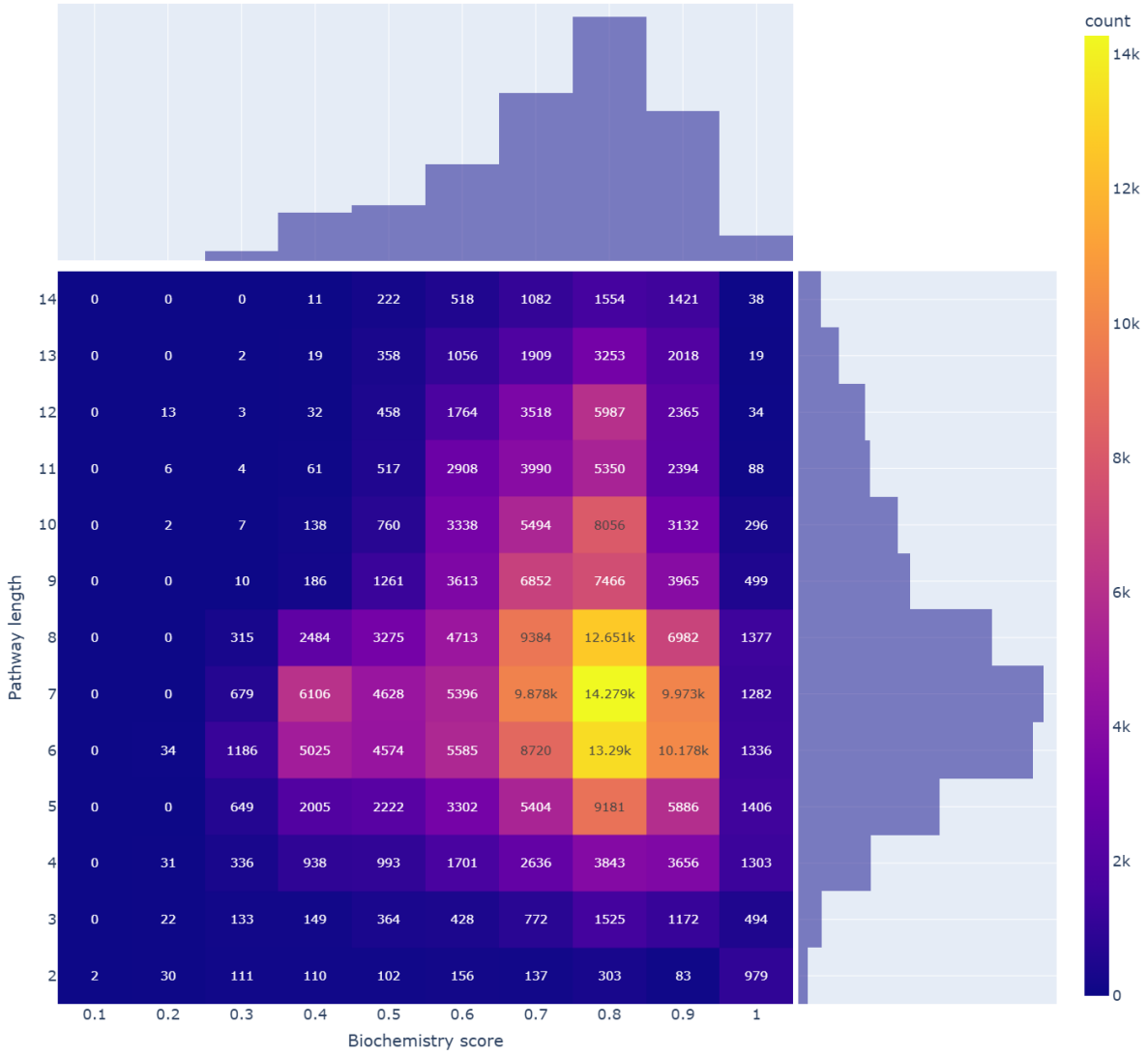


Figure 45: Heatmap representing the pathway lengths and biochemical scores of the designed carbon-carrying pathways.

6.4 Summary

In this chapter, I have identified a set of candidate carbon- and nitrogen-exchange compounds in a consortium of the diazotroph *A. vinelandii* and the photoautotroph *Synechocystis* 6803. Candidates were selected from the MetaNetX database (Ganter et al., 2013; Moretti et al., 2016a; Moretti et al., 2016b; Moretti et al., 2021) and the RetroPathRL bioretrosynthesis algorithm (Koch et al., 2019) was used to identify which candidate compounds carrying each nutrients could be synthesised in the donor chassis – *A. vinelandii* for nitrogen and *Synechocystis* 6803 for carbon – and broken down in the recipient chassis. This approach has led to the identification of 3,693 candidate carbon-exchange compounds with a total 294,331 possible exchange pathways, and 1,488 candidate nitrogen-exchange compounds in 139,139 possible exchange pathways.

One issue encountered when implementing predicted pathways in this chapter is data loss when interconverting between different compound representations. While all compounds were retrieved from in InChI format to ensure unambiguous representation of molecules, outputs from RetroPathRL consist of a combination of InChI, InChIKeys (InChI Trust, 2024) and reaction SMILES (Leach et al., 1999). Care must be taken to when interconverting these formats, especially as reaction SMILES[2] may not preserve stoichiometry and charge information. Additionally, SMILES often do not contain explicit declaration of hydrogen atoms. Stoichiometry plays an especially important role in biochemical reactions, as different isomers cannot participate in the same catalytic pathways in the same way. Charge information can also be lost during the pipeline. While InChI and, to a lesser extent, InChIKey both represent molecular charge, this information can be lost during compound sanitisation which takes place during a RetroPathRL run (Koch et al., 2019; Laudrum & others, 2024). One major

[2] Isomeric SMILES are another format that may be used but was not implemented in this case.

disadvantage of RetroPathRL, the bioretrosynthesis software employed in this chapter, is its lack of stereoawareness, meaning that chemical transformations encoded within do not take stereochemistry into account. Both these limitations must be taken into consideration when implementing predicted pathways *in vitro*.

Chapter 7
Implementation and Evaluation of the Nutrient-Exchange Potential of Bioretrosynthetically Identified Pathways

7.1 Introduction

7.1.1 Rationale

A genome-scale metabolic model developed in Section 3.3 was used to test the feasibility of novel pathways designed using bioretrosynthesis in Chapter 6. Constraint-based analysis using a GEM is a good tool to enable the availability of precursors, cofactors and energy balances required for a novel pathways in a given chassis, hereby evaluating whether a novel pathway is feasible in the context of the chassis of choice. In the case of nutrient-exchange pathways, FBA not only enables the qualitative assessment and quantification of potential compound yields, but also enables us to evaluate whether a compound provided by the donor strain can drive sufficient biomass flux to be compatible with survival.

In this chapter, biosynthesis and breakdown pathways for nutrient exchange compounds identified in Section 5.3.7 are implemented and tested in a genome-scale metabolic model of a consortium of the diazotroph *A. vinelandii* and the photoautotroph *Synechocystis* 6803 constructed in Section 3.3 of this thesis.

7.1.2 Aims

- Implement stoichiometrically balanced biosynthesis and breakdown pathways in GEM-compatible format
- Determine maximum yield for each compound

- Test viability of *Synechocystis* 6803 with various nitrogen-exchange pathways under nitrogen limitation
- Test viability of *A. vinelandii* with various carbon-exchange pathways under carbon limitation
- Test viability of whole consortium with various combinations of carbon- and nitrogen-exchange pathways under carbon and nitrogen limitation
- Evaluate the feasibility of implementing pathways predicted to support best growth rates *in vivo*

7.2 Implementation of Nutrient Exchange Pathways in Metabolic Models

7.2.1 Compound Renaming for Compatibility with Cobra Toolbox 3.0

Compound representations used in the output from bioretrosynthetic pathway prediction (InChI) had to be converted to match identifiers used in GEM (BiGG metabolite names). A conversion table matching InChI and BiGG identifiers created from MetaNetX and BiGG database entries (prepared in Materials and Methods, Section 2.3.2) was used to carry out conversions. For some InChI identifiers, a metabolite with differing protonation state had to be used, as a perfect match was unavailable. In this case, I am working under the assumption of spontaneous reversible protonation and deprotonation. For some compounds, a conversion was not possible as no BiGG identifier was found to match any protonation form of the compound – these compounds are ones which had not previously appeared in any GEM deposited in the BiGG database. These novel metabolites were represented using their InChIKey. InChIKey was chosen over InChI in this application due to its shorter length; InChIKeys are fixed at 27 characters, while InChIs can be of unlimited length, growing with compound molecular weight and complexity (InChI Trust, 2024). A shorter variable length decreases the memory requirements for each script and improves code legibility to improve ease of debugging where necessary.

7.2.2 Stoichiometric Balancing

Pathways identified through retrosynthesis were modified to account for correct stoichiometry, as RetroPathRL does not account for atomic balance (Koch et al., 2019) using a

python script. Briefly, a set of linear equations was solved to balance atom counts in reactants and products. Where necessary, additional water molecules, hydrogen ions, nucleotide tri- and diphosphates or CoA where applicable were added to balance reactions. The Brenda protein database was used to identify appropriate hydrogen-carrying cofactors (NADH, NAPH, FAD) (Chang et al., 2021), which were added in place of hydrogen ions only where appropriate. Refer to Materials and Methods, Section 2.3.6 for detail. Only 10,695 of 83,144 (12.86%) reactions in nitrogen-carrying pathways and 48,282 of 422,324 (11.43%) reactions in carbon-carrying pathways could be balanced programmatically, as only a set of small additional molecules required for balancing could be accounted for in an automated process. Manual balancing would be necessary in cases where large fragments were missing but was not pursued as balancing the total set of 42,592 was deemed temporally intractable – assuming 5 minutes per pathway, this task alone would require 3,549 hours.

7.2.3 Implementation

Reactions were added to the GEM for the *A. vinelandii* – *Synechocystis* 6803 consortium using the `addReaction()` function in Cobra Toolbox 3.0 (Heirendt et al., 2019) using BiGG or InChIKey metabolite identifiers. Where possible, the balanced equation was used. All new reactions were added to the model unconstrained, with reaction flux bounds from -1000 to +1000 mmol/gDWh. Prior to implementation, the BiGG ID of reactants at the first step within a biosynthetic pathway or products at the last step in a breakdown pathway were double checked to ensure that the pathway is connected to the native metabolism present in the model to avoid orphan reactions which could carry no flux.

7.3 Maximum Biosynthetic Potential

A total set of 8,384 biosynthetic pathways for nitrogen-containing compounds in *A. vinelandii*, and 34,208 biosynthetic pathways identified in Chapter 6 for carbon-containing compounds in *Synechocystis* 6803 was individually implemented in the respective member of the consortium using a version of the model simulating rich media conditions to explore the maximum biosynthetic potential of each species in this medium configuration. Only such compounds which can be synthesised in sufficient quantities can be expected to function as useful nutrient-exchange compounds. Additionally, this set of results provides an exploration of the biosynthetic capabilities of both organisms, upon which future research can build when using *A. vinelandii* or *Synechocystis* 6803 as a bioproduction chassis.

Flux through the final reaction of each biosynthetic pathway when optimised for this flux in FBA was used as proxy to evaluate the maximum achievable product yield. First, the pathways were evaluated in the context of the full consortium metabolic model as developed in Section 3.3 under rich media (carbon and nitrogen-replete) conditions. This set of experiments was then also carried out in conditions where diazotrophy and photoautotrophy-specific metabolic modifications were removed – namely, removing the constraint on *Synechocystis* 6803 oxygen-sensitive hydrogenase (Mcintosh et al., 2011) and *A. vinelandii* ATP maintenance requirement (Alleman et al., 2021). Note that no nutrient exchange between cells was implemented at this stage.

Interestingly, less than 4% of *Synechocystis* 6803 pathways and less than 0.5% of *A. vinelandii* pathways could reach a non-zero maximum pathway flux when optimised for (Table 44, Figure 46). Counterintuitively, fewer compounds and pathways show a non-zero maximum pathway flux when optimised for in both unconstrained models (Figure 47, Figure 48) – a model with fewer constraints would normally be expected to produce a larger set of feasible

solutions. In this case both unconstrained models were originally prepared under the assumption that low numbers of positive non-zero maximum pathway fluxes in the original diazotrophy/photoautotrophy-constrained models may be the result of overconstraining the model in the first place; this appears not to be the case. As orphan reactions had also been controlled for during implementation (Section 7.2.3), this result indicates that pathways with a zero maximum flux reflect that nutrient demands – fixed carbon, nitrogen, or cofactors – required to drive flux are not being met. Hence, while a large number of pathways is theoretically engineerable, only few support flux without additional metabolic pathway engineering. For pathways with non-zero maximum pathway flux in both conditions, there was a very high degree of correlation between predictions of photoautotrophy/diazotrophy-constrained and unconstrained models in both consortia members (Figure 47, Figure 48).

Table 44: Numbers of compounds and pathways with a positive non-zero maximum pathway flux when optimised for.

	***Synechocystis* 6803**	***A. vinelandii***
Diazotrophy/photoautotrophy-based constraints	**176**/4694 compounds **905**/34229 pathways	**27**/1373 compounds **40**/8384 pathways
Unconstrained	**175**/4694 compounds **961**/34229 pathways	**12**/1373 compounds **16**/8384 pathways

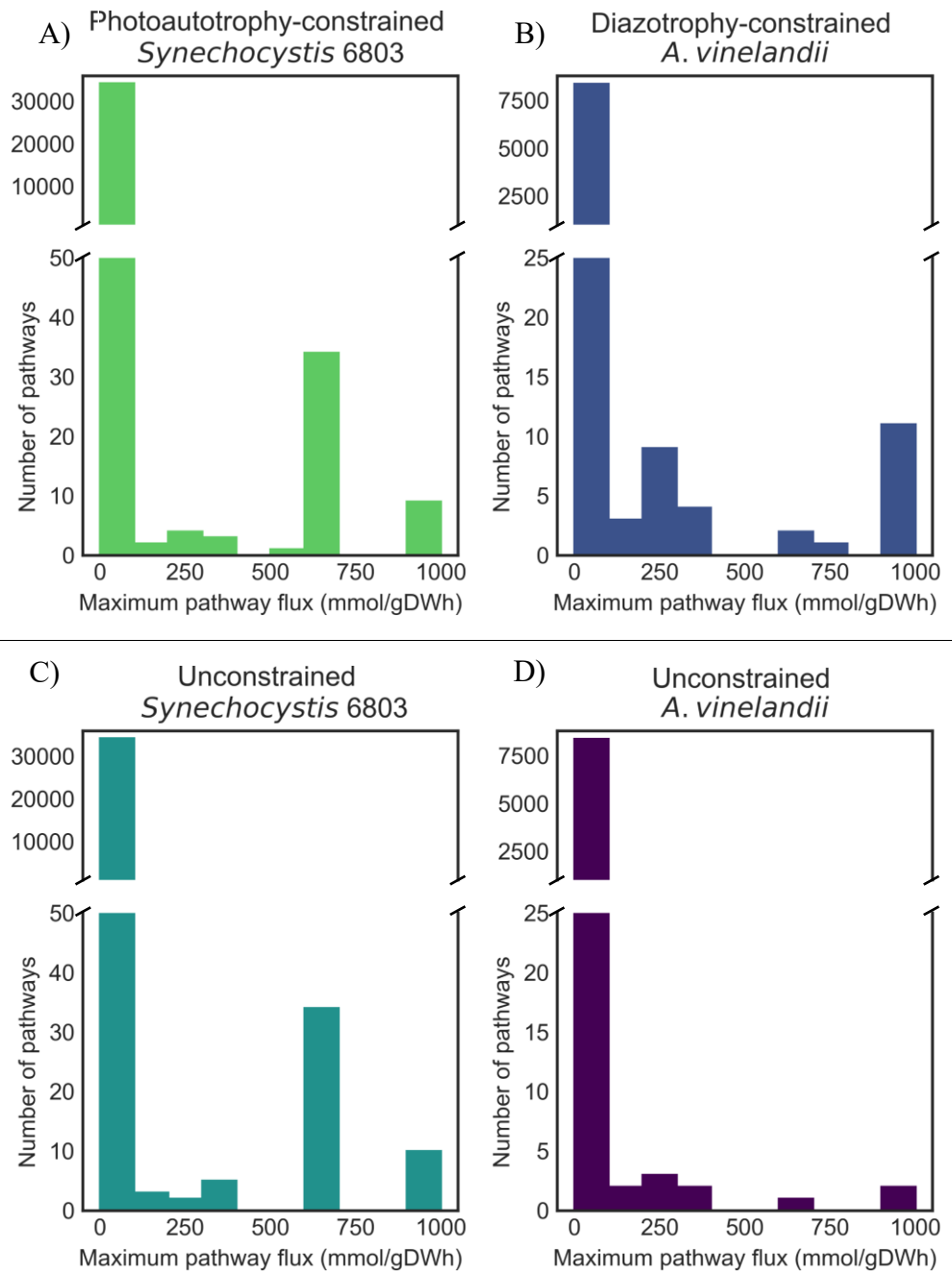


Figure 46: Histograms showing a distribution of maximum pathway flux (mmol/gDWh) through final reaction within each biosynthetic pathway in rich media in A) *Synechocystis* 6803 constrained to reflect photoautotrophic conditions, B) *A. vinelandii* constrained to reflect diazotrophic conditions, C) *Synechocystis* 6803 with free hydrogenase flux, and D) *A. vinelandii* unconstrained for ATP maintenance.

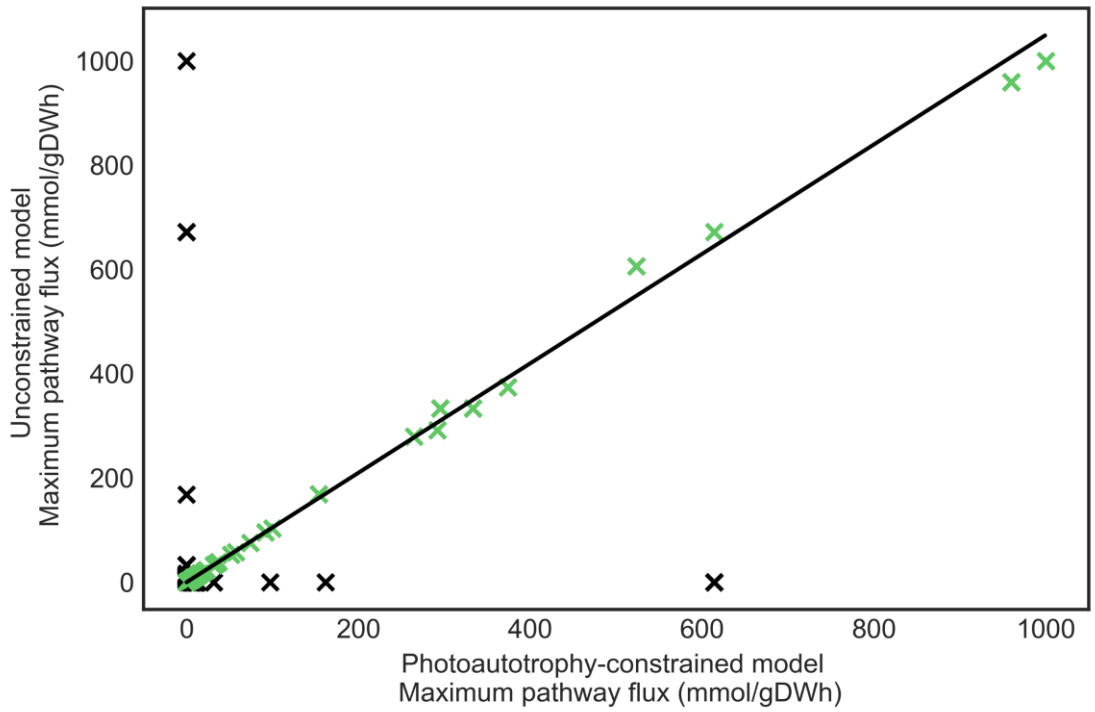


Figure 47: Correlation of predicted maximum pathway flux (mmol/gDWh) as proxy for maximum compound yield for compounds in *Synechocystis* 6803 in rich media with photoautotrophy-specific constraints and *Synechocystis* 6803 in rich media unconstrained for photoautotrophy. Each marker represents a single pathway. Linear regression calculated using only pathways for which maximum predicted flux is positive in both constrained and unconstrained conditions (green markers) shows a correlation coefficient $R^2 = 0.998$.

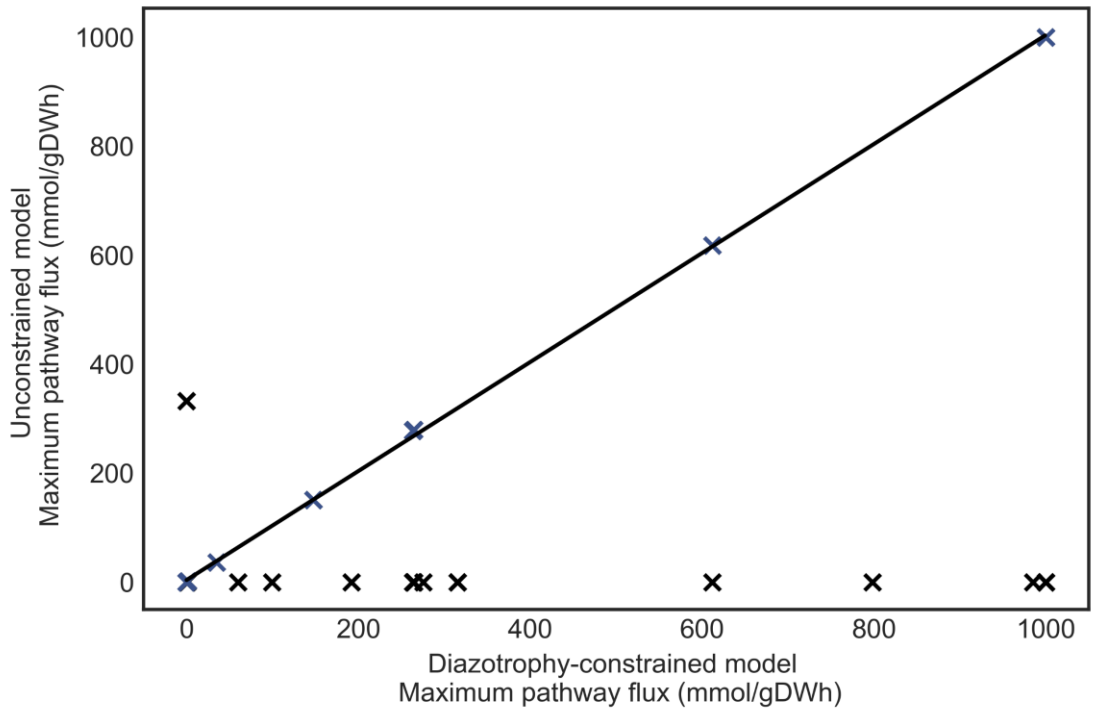


Figure 48: Correlation of predicted maximum pathway flux (mmol/gDWh) as proxy for maximum compound yield for compounds in *A. vinelandii* in rich media with diazotrophy-specific constraints and *A. vinelandii* in rich media unconstrained for photoautotrophy. Each marker represents a single pathway. Linear regression calculated using only pathways for which maximum predicted flux is positive in both constrained and unconstrained conditions (violet markers) shows a correlation coefficient $R^2 = 0.999$.

7.4 Growth Rates of Consortia with Nutrient-Exchange Pathways

7.4.1 Pathways in Media Lacking One Nutrient at a Time

Nutrient exchange pathways predicted with bioretrosynthesis in Chapter 6 were implemented into GEM model of *A. vinelandii*-*Synechocystis* 6803 consortium. First, a total set of 139,139 nitrogen-exchange pathways was implemented in a model constrained to reflect nitrogen-deplete conditions. Separately, a subset of 100,000 of the total 294,331 carbon-exchange pathways was implemented in a model constrained to reflect carbon-deplete conditions. In the case of carbon-exchange compounds, only a subset could be trialed due to limited capacity on the computing cluster; pathways were prioritised based on highest scores per Section 6.3.3.

In both conditions, biomass objective functions of both consortium partners were individually optimised for in every pathway condition. Results were subsequently subset to exclude all carbon-exchange pathways that could not support growth of *A. vinelandii* in carbon-deplete conditions and all nitrogen-exchange pathways that could not support growth of *Synechocystis* 6803 in nitrogen-deplete conditions; therefore, only effective nutrient-exchange pathways were preserved for further analysis.

36 nitrogen-carrying pathways capable of supporting the growth of *Synechocystis* 6803 in nitrogen-deplete media were identified with *Synechocystis* 6803 growth rates ranging from 0.0005 gDW/h to 0.022 gDW/h; in all 36 pathways, the maximum *A. vinelandii* growth rates was found to equal 0.0034 gDW/h (Figure 49). For context, these growth rates lie within a similar range as those reported for rationally designed pathways in Chapter 5 (Figure 39, Figure 40). 447 carbon-carrying pathways capable of supporting the growth of *A. vinelandii* in carbon-

deplete media were identified. *A. vinelandii* growth rates range between 0.0001 gDW/h and 7.2349 gDW/h, and *Synechocystis* 6803 growth rates range between 0.0004 gDW/h and 0.4252 gDW/h (Figure 49). While a higher efficiency of nutrient transfer has previously been demonstrated to predict higher growth rates, scepticism is warranted. It is possible that these "super-efficient" strains may be the result of model artifacts such as futile energy-generating cycles, rather than being reflective a real phenomenon; this would be consistent to these pathways corresponding to compounds for which maximum biosynthetic flux was found to be unconstrained at 1000 mmol/gDWh. Nevertheless, all 36 nitrogen-carrying pathways and 447 carbon-carrying pathways were included in further analysis.

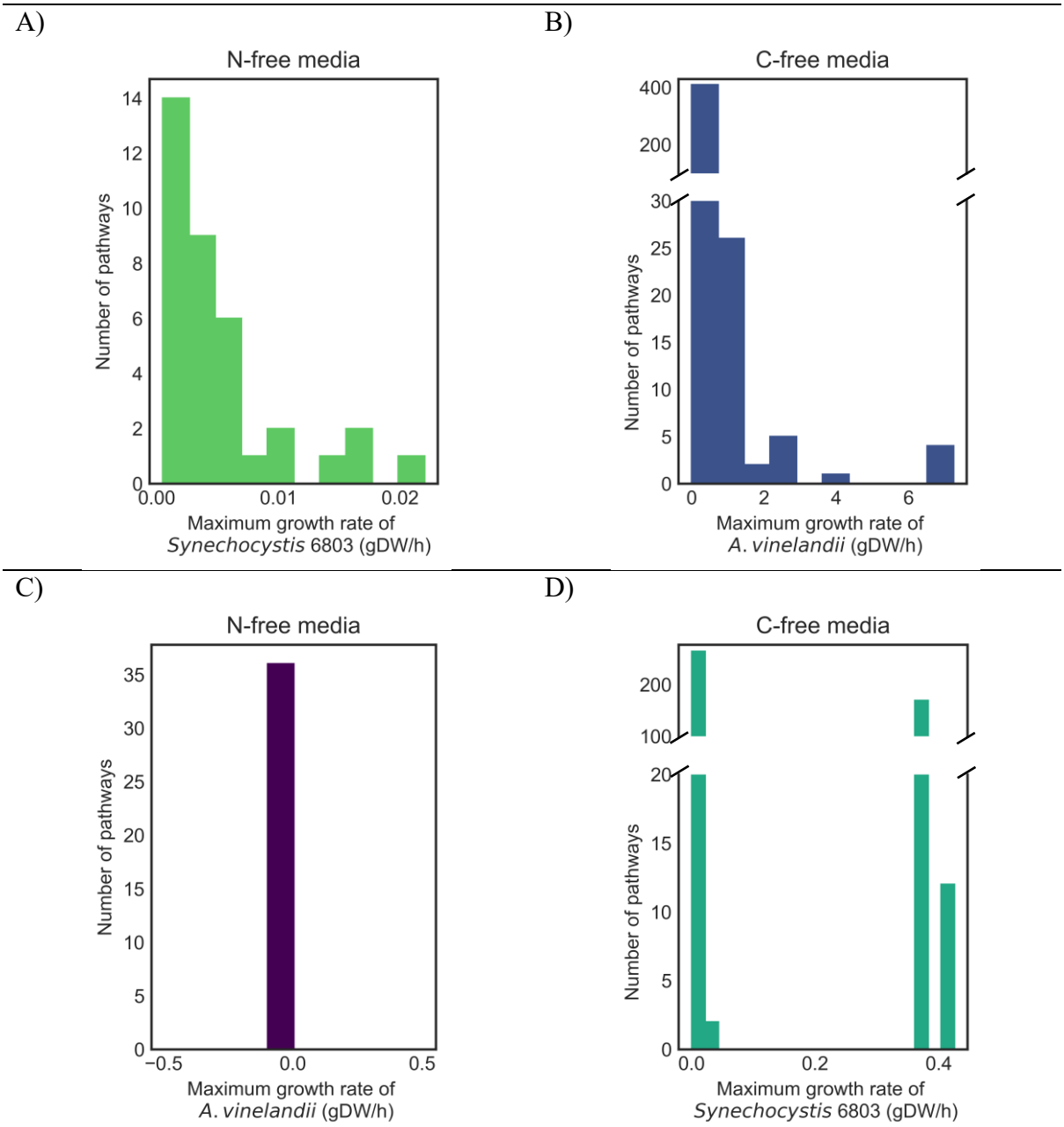


Figure 49: Histograms representing distribution of maximum growth rate (calculated as biomass objective function flux gDW/h) in A) *Synechocystis* 6803 in consortium in nitrogen-free media and B) *A. vinelandii* in consortium in carbon free media, C) *A. vinelandii* in consortium in nitrogen-free media, and D) *Synechocystis* 6803 in consortium in carbon-free media

7.4.2 Pathways Evaluated in Media Lacking Both Fixed Carbon and Fixed Nitrogen

A combinatorial approach was used to investigate all combinations of nitrogen- and carbon carrying pathways which showed positive maximum biomass objective function flux in media lacking one nutrient at a time for the species auxotrophic for a given nutrient (Section 7.4.1). This resulted in a total 16,092 pathways, which were implemented in a consortium model reflecting media deplete in both carbon and nitrogen – only such pathway combinations compatible with the provision of both essential nutrients to both partners would result in a positive biomass objective function flux for both.

Of the 16,092 pathways investigated, 11,696 were found to enable non-zero growth rates for *A. vinelandii* and *Synechocystis* 6803 (Figure 50, Figure 51, Figure 52). All 36 nitrogen-carrying and 447 carbon-carrying pathways identified earlier are represented. Growth rates predicted for *Synechocystis* 6803 range from 0.0001 to 1.47 gDW/h, and growth rates for *A. vinelandii* range from 0.0001 to 7.3021 gDW/h. The median and interquartile range of the growth rates of both organisms fall within a similar range as growths observed in previous *in vitro* and *in silico* experiments in this thesis; growth rates at the higher ends of the range for *A. vinelandii*, however, fall outside. Notably, the heatmap representation in Figure 52 illustrates that highest *A. vinelandii* growth rates are associated with only a few carbon-carrying pathways across a variety of nitrogen-carrying pathways. The pathways with the 25 highest *A. vinelandii* growth rates all use the same carbon-carrying pathway, and only 4 different carbon-carrying pathways are responsible for the top 100 *A. vinelandii* growth rates out of a total 447 possible pathways in the dataset (<1% of the pathways). In contrast, the top 100 *Synechocystis* 6803 growth rates are associated with 35 different pathways of a total 36 pathways implemented.

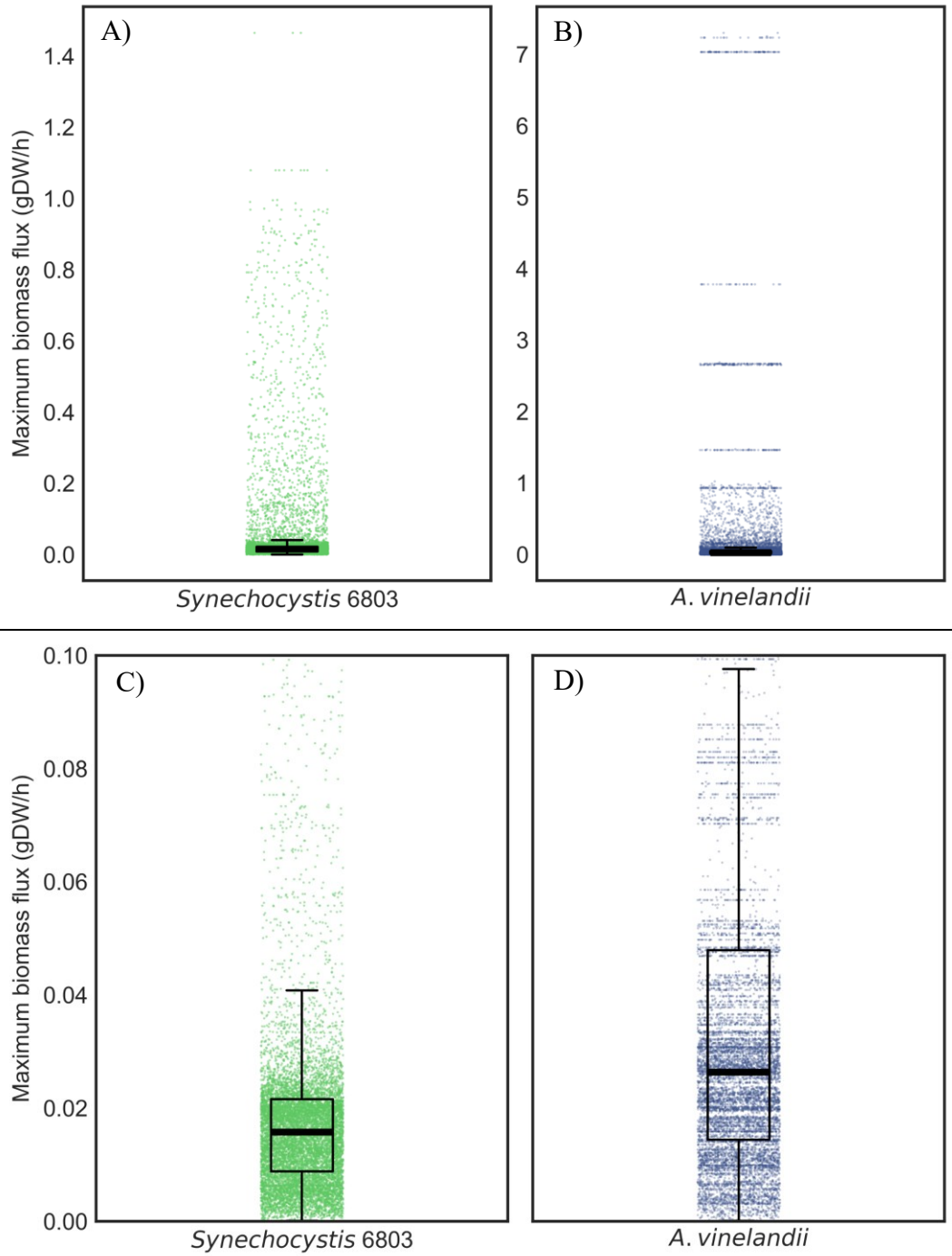


Figure 50: Distribution of the maximum biomass flux for A, C) *Synechocystis* 6803 in nitrogen-deplete media and B, D) *A. vinelandii* in carbon-deplete media in consortium with nutrient-exchange pathways for the respective essential nutrient. Figures C and D represent zoomed-in version of A and B for visual clarity. Box plots overlaid over individual datapoints extend from the first quartile to the third quartile of the data, with a line on the median. Whiskers extend to the farthest point within 1.5x the interquartile range from the box.

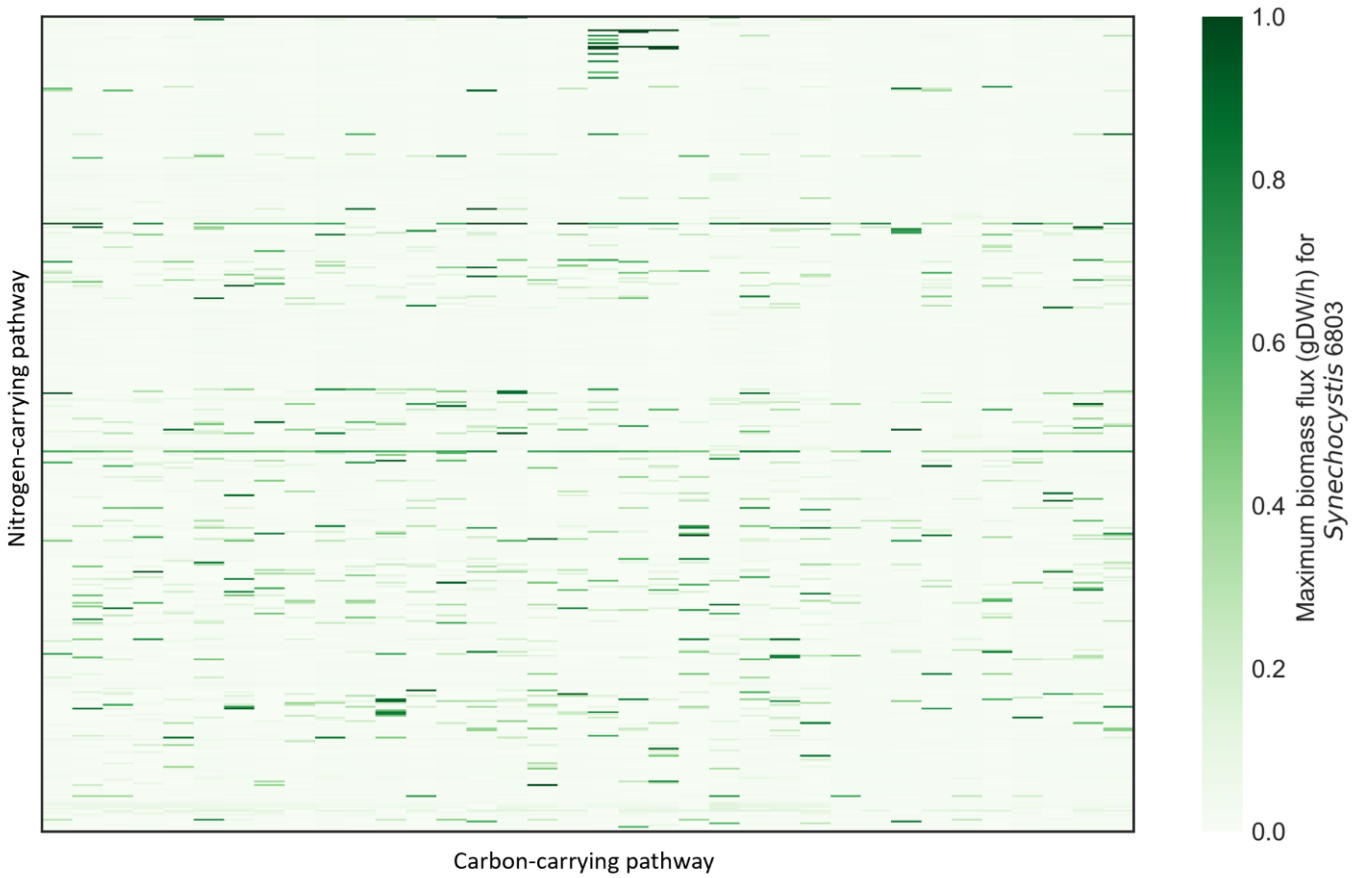


Figure 51: A heatmap representing the growth rates of *Synechocystis* 6803 in carbon-free, nitrogen-free media representing a combination of 36 nitrogen-carrying and 447 carbon-carrying pathways in the consortium. Only pathways with a positive maximum growth rate for both species are represented. Darker shades represent higher growth rates. Colour scale is trimmed at an upper limit of 1 for higher visual clarity; for the full range of growth rates, refer to Figure 50A.

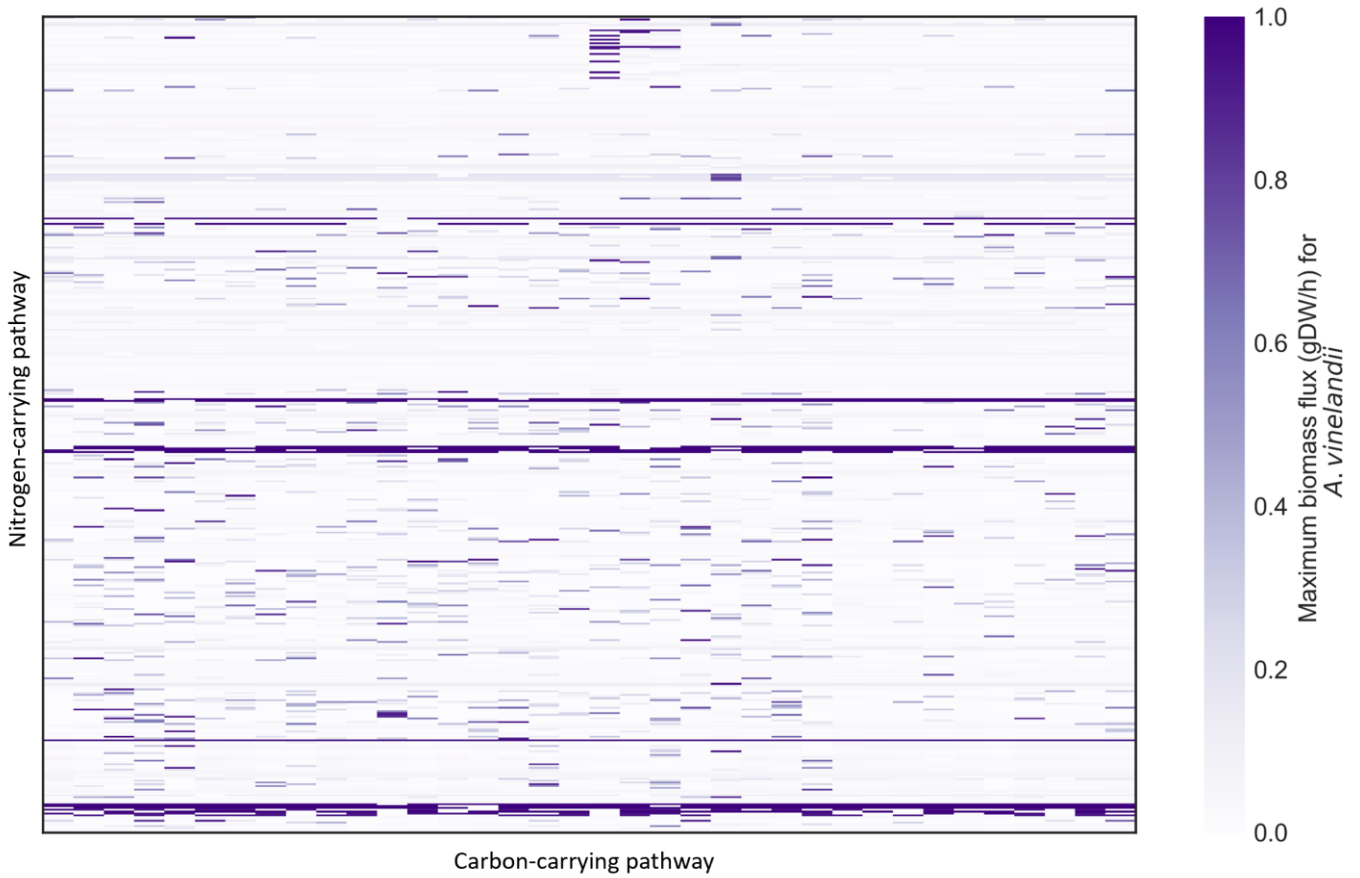


Figure 52: A heatmap representing the growth rates of *A. vinelandii* in carbon-free, nitrogen-free media representing a combination of 36 nitrogen-carrying and 447 carbon-carrying pathways in the consortium. Only pathways with a positive maximum growth rate for both species are represented. Darker shades represent higher growth rates. Colour scale is trimmed at an upper limit of 1 for higher visual clarity; for the full range of growth rates, refer to Figure 50B.

7.5 Summary

In this chapter, candidate nutrient-exchange pathways identified through bioretrosynthesis (Chapter 6) were implemented into a genome-scale metabolic model of a consortium of *A. vinelandii* and *Synechocystis* 6803. The maximum biosynthesis flux in rich media was modelled. Subsequently maximum biomass flux of consortium partners in media lacking only nitrogen or only carbon at a time was determined. 36 nitrogen-carrying and 447 carbon-carrying pathways found to carry non-zero flux in these conditions were finally combinatorically implemented in a model reflecting simultaneously carbon and nitrogen-deplete conditions, yielding 11,696 pathways with non-zero growth rates for both organisms.

Chapter 8
Conclusion and Discussion

This thesis contributes to the establishment of an obligate nutrient-exchange consortium between *A. vinelandii*, a diazotroph, and *Synechocystis 6803*, a photoautotroph directed at developing more efficient transfer of fixed carbon and nitrogen between plants and free-living rhizosphere diazotrophs. Investigating alternative metabolic symbioses forms a part of addressing one of today's greatest biotechnological challenges: humanity's overreliance on synthetic nitrogenous fertilisers produced using the Haber-Bosch process. We propose that using metabolic engineering to optimise soil microbiota for efficient metabolic exchange with plants can be used to improve the nitrogen balance of agriculture in the future.

8.1 Establishing a Consortium with Nature-Mimicking Nutrient Exchange

A consortium of the fully aerobic rhizosphere diazotroph *Azotobacter vinelandii* and the photoautotroph non-diazotrophic cyanobacterium *Synechocystis* 6803 was established. A combination medium, BNxBG11, was developed to support the growth of both organisms. A novel physical barrier system was introduced to enable the quantification of the growth of each organism while enabling nutrient exchange between partners. Though user-friendly, the system requires further validation to confirm nutrient equilibration across compartments. This work demonstrates a straightforward strategy for high-throughput viability and tolerance screening in novel cocultures.

The naturally occurring nutrient exchange system in the multicellular diazotrophic and photoautotrophic cyanobacterium *Anabaena sp.* PCC 7120 was used as a guide to engineering a proof-of-principle nature-mimicking nutrient-exchange system in the *A. vinelandii*-

Synechocystis 6803 consortium. Genetic parts required for the setup of a GS/GOGAT nutrient exchange system were designed, synthesised and cloned. Genetic cassettes for the deletion of the nitrogenase activity regulator (*nifL*) and glutamate synthase (*GOGAT*) were assembled and transformed into *A. vinelandii*; further work is required to confirm genomic integration. For *Synechocystis 6803*, cloning and assembly of *ybjL* glutamate transporter and *cscB* sucrose-proton symporter were initiated. Additionally, a HPLC protocol for the quantification of glutamate and glutamine using OPA derivatisation was developed.

The work carried out here provides a foundation to setting up and evaluating a nutrient exchange consortium between a photoautotroph and diazotroph. Unfortunately, portions of this experimental section could not be completed as I had to halt all wet lab research due to a sudden onset of illness preventing safe return to lab work. Masters and undergraduate research projects were subsequently proposed to carry on this work, these projects were unfortunately without applicants. To evaluate the performance of this consortium once all engineering steps are complete, a follow-up project will need to culture the consortium in media lacking carbon and nitrogen supplementation. The high rate of respiration characteristic for *A. vinelandii* is a likely challenge to the final success of this consortium. *A. vinelandii* must consume a very high quantity of fixed carbon to not only provide enough ATP for the costly process of nitrogen fixation, but also to support sufficient oxygen scavenging enabling the microaerobic environment vital to nitrogenase activity. In a downstream project, glutamate auxotrophy can be used to further fortify the dependency of *A. vinelandii* on *Synechocystis* 6803 as a useful tool to improve selection pressure in any downstream adaptive laboratory evolution (ALE) pursuits. This work could provide additional insights into the gene regulatory environment surrounding the engineered pathway.

8.2 *In silico* Reconstruction of *A. vinelandii-Synechocystis* 6803 Consortium

I have developed a two-cell *in silico* metabolic model of the photoautotroph-diazotroph consortium based on previously published genome-scale reconstructions of *A. vinelandii* and *Synechocystis* 6803. The model was designed to represent nutrient exchange between the organisms and constrained to reflect experimental media conditions. Flux balance analysis (FBA) was used to confirm qualitative alignment between predicted and experimentally observed growth rates, although predicted growth rates were overestimated relative to *in vitro* measurements.

Genome-scale metabolic modelling plays an important role in metabolic engineering, allowing for the evaluation of an engineered pathway within the metabolic context of a given chassis. Unfortunately, there are several limitations associated with GEMs. Firstly, currently available models are incomplete, and considerable genome annotation work accompanied by fluxomics is required to improve GEM coverage. In 2014, for example, only 33% of the enzymatic activities in the BRENDA database were found to exist within published GEMs. However, the number of published GEMs and the number of reactions and metabolites in them are continually growing (Monk et al., 2014) and updated versions of the same model are regularly published; and it is likely only a question of time until GEM coverage approaches completion, at least for less exotic workhorse organisms. While model coverage is improving, the refinement of appropriate model constraints is likely to remain as a major challenge in GEMs. Recently, Ai et al. (2024) have proposed an elegant approach to the iterative refinement of GEMs using the design-build-test-learn framework popular in synthetic biology, which was originally adapted from engineering disciplines. The authors have implemented a novel approach called Boolean Matrix Logic Programming (BMLP) to decrease the experimental

design space required for *in vitro* experiments enabling rapid optimisation of GEMs. While these limitation underscore the need to consider discrepancies arising from reaction kinetics, gene regulation, and environmental conditions when translating model predictions to laboratory settings, GEMs remain a useful tool for guiding strain engineering and pathway optimization.

8.2.1 Retrosynthesis and a High-Throughput Unbiased Approach to *in silico* Metabolic Engineering

A high-throughput approach was employed in the screening of candidate nutrient-exchange compounds and pathways in which these are utilised. A total set of 11,696 carbon-nitrogen exchange pathways capable of supporting the growth of both consortium partners was identified using candidate compounds deposited in the MetaNetX database, and an additional 2,871 exchange pathways were implemented using metabolites synthesised by the wild-type consortium.

Clustering based on the chemical similarity of exchange compounds was used to investigate the relationship between exchange compound structure and maximum biomass objective flux achievable across all pathways utilising it (Figure 53). Briefly, exchange compounds in promising nutrient-exchange pathways were converted from their InChI representation to Morgan Fingerprints, in which the 2D conformation of atoms in a molecule is encoded, using the RDKit Python package (Laudrum & others, 2024; Rogers & Hahn, 2010). Uniform Manifold Approximation and Projection (UMAP) (Healy & Mcinnes, 2024; McInnes et al., 2018) and the associated python package (https://github.com/lmcinnes/umap) was used for clustering and dimensionality reduction. Tanimoto (Jaccard) similarity, which has been described earlier in Section 6.3.1.1., was used for clustering of compounds. The Tanimoto similarity of two compounds is defined as the intersection of two molecular fingerprints over

their union; compounds with more shared features therefore have a higher Tanimoto similarity. This analysis includes both compounds from pathways identified using bioretrosynthesis (Sections 6.3 and 7.4.2) and compounds derived from compounds occurring in the wild-type consortium model (Section 5.2.1).

While distinct clusters representing groups of molecules with higher structure similarities were successfully identified using this approach, no relationship was found between the cluster to which a compound belongs and the maximum biomass objective flux (gDW/h) achievable in either *A. vinelandii* or *Synechocystis* 6803 if a given molecule is used (Figure 53). This indicates that the chemical identity of a compound does not correlate to its suitability as a nutrient-exchange compound, highlighting the advantages of a high-throughput unbiased approach to pathway development. In future work, this clustering approach could be extended to contain all compounds all 1,048,228 carbon-containing and 381,616 nitrogen-containing compounds from the MetaNetX database that had been initially investigated; this clustering was however not pursued as part of this project due to added computational complexity.

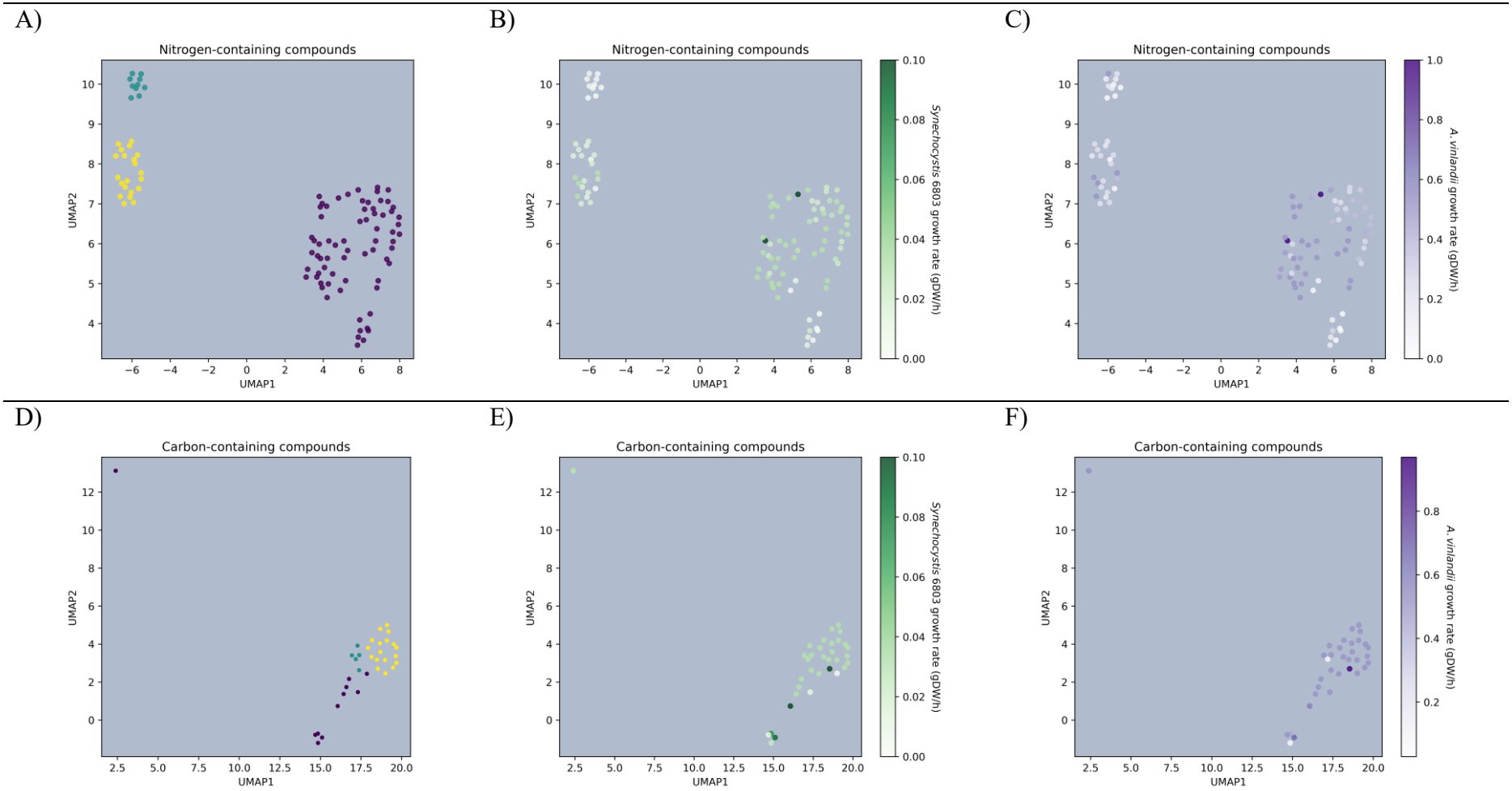


Figure 53: UMAP/hdbscan clustering analysis of A-C) nitrogen carrying compounds and D-F) carbon-carrying compounds in media colored by A, D) cluster, B, E) maximum *Synechocystis* 6803 biomass flux (gDW/h) and C, F) maximum *A. vinelandii* biomass flux (gDW/h).

Additionally, the distribution of maximum growth rates predicted for nutrient-exchange pathways designed using bioretrosynthesis was compared to the distribution of maximum growth rates in exchanges derived from the set of nitrogen and carbon containing compounds already present in the wild-type (Figure 54).

To recapitulate, in the case of pathways utilising native metabolites, *Synechocystis* 6803 growth rates ranged up to 3.60×10^{-2} gDW/h (Figure 31) and *A. vinelandii* growth rates up to 5.18×10^{-1} gDW/h (Figure 32). In bioretrosynthetically designed new-to-nature pathways, growth rates predicted for *Synechocystis* 6803 reach up to 1.47 gDW/h (Figure 51), and growth rates for *A. vinelandii* to 7.3021 gDW/h (Figure 52). While ranges of growth rates achieved in "new" and "native"-like pathways vary considerably, the highest density of pathways is observed around the same values.

Table 45: Mean, lower quartile (Q1) and upper quartile (Q3) of maximum growth rates (gDW/h) predicted for *Synechocystis* 6803 and *A. vinelandii* in consortium in media lacking both nitrogen and carbon between pathways designed using bioretrosynthesis (new) and derived from the set of nitrogen and carbon containing compounds already present in the wild-type (native).

		native	**new**
***Synechocystis* 6803**	**mean**	1.17×10^{-2}	4.27×10^{-2}
	Q1	0	8.80×10^{-3}
	Q3	2.12×10^{-2}	2.16×10^{-2}
A. vinelandii	**mean**	1.96×10^{-1}	1.94×10^{-1}
	Q1	0	1.44×10^{-2}
	Q3	3.51×10^{-2}	4.79×10^{-2}

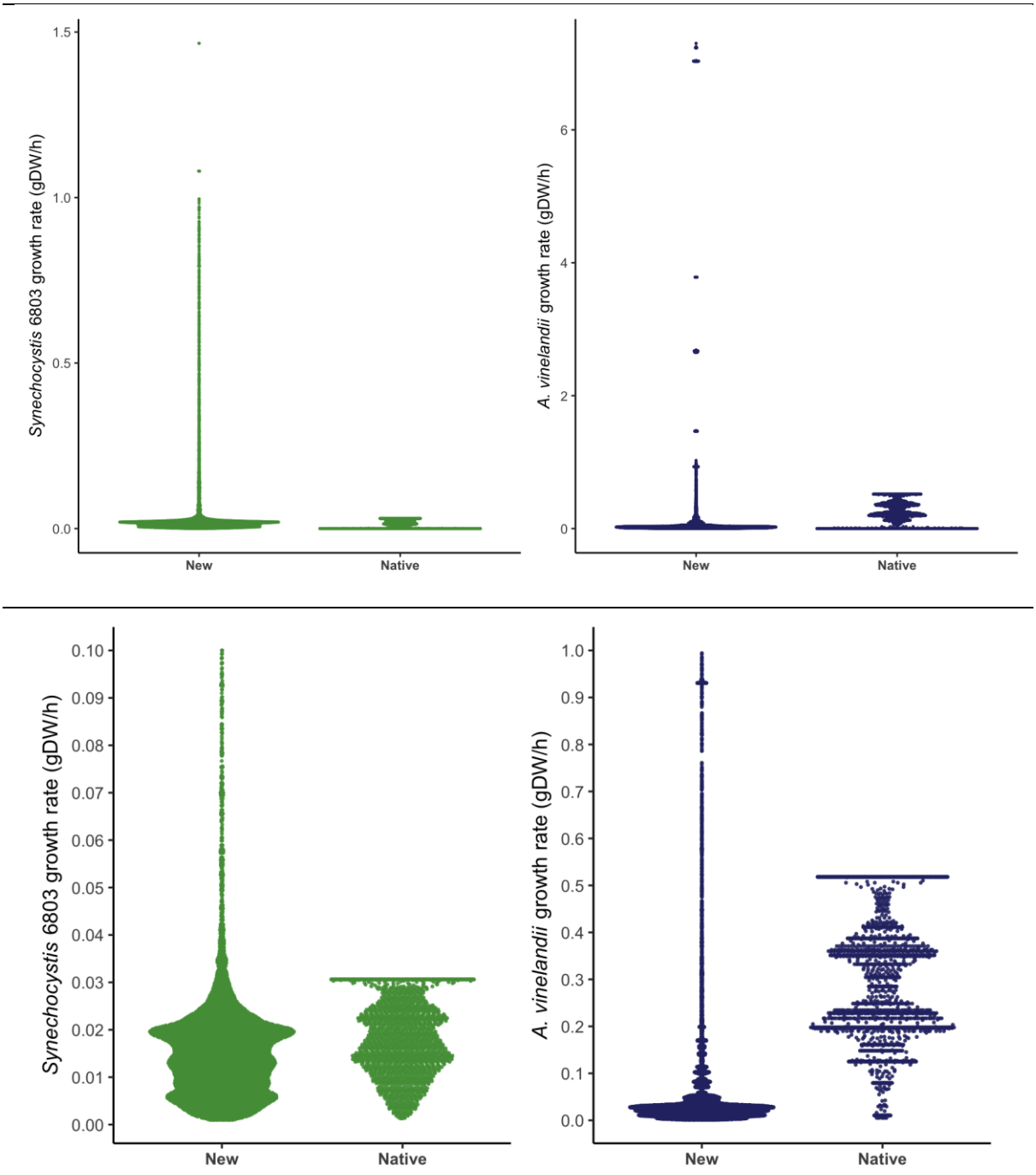


Figure 54: Comparison of the distribution of maximum growth rates predicted for *Synechocystis* 6803 (left) and *A. vinelandii* (right) in consortium in media lacking both nitrogen and carbon between pathways designed using bioretrosynthesis (new) and derived from the set of nitrogen and carbon containing compounds already present in the wild-type (native). The top graphs represent a zoomed-out version of graphs containing all datapoints including pathways showing zero growth, bottom two graphs are presented for visual clarity.

In the case of *A. vinelandii*, nutrient exchanges employing native metabolites appear to be much more reliable than those with novel compounds; this phenomenon is not observed in the growth rates of *Synechocystis* 6803. In the case of both organisms, however, new-to-nature exchange pathways reveal some candidates predicting growth rates considerably higher than those observed in exchanges utilising native compounds. This highlights one of the advantages of a high-throughput approach to candidate discovery: more extreme values can be discovered. This includes extremes in the negative, as only a small proportion of candidate exchange compounds has yielded a promising nutrient exchange pathway at the end of the pipeline presented in this thesis, but also in the positive, with fluxes beyond those identified using a set of native compounds, or using rationally-designed nutrient exchange pathways in Section 5.3.

Bioretrosynthesis is a very useful tool for computer-aided design in metabolic engineering, enabling efficient transversal of the design space. Over the past decade, bioretrosynthesis has been used in the construction of many metabolic pathways, largely for the synthesis of natural products in a non-native chassis, sometimes including a final bioconversion step not occurring in nature (Li et al., 2018; Curran et al., 2013). Full computer-aided bioretrosynthesis of a xenobiotic is still exceedingly rare (Yim et al., 2011; Lin, Warden-Rothman & Voigt, 2019). Gricourt et al. (2024) have highlighted that published bioretrosynthesis tools in their current state are well-suited to exploratory analysis in large datasets, consistent with the use case presented in this thesis. Recent developments in artificial intelligence and machine learning show promise for future bioretrosynthesis algorithms, with AI strategies improving the efficiency of combinatorial graph search. Additionally, string-based molecular representations like the InChI identifier are compatible with AI tools for natural language processing, which is likely to become a key cheminformatics tool in the years to come. Methods for retrobiosynthesis with a focus on artificial intelligence have recently been reviewed by Gricourt et al. (2024).

Extensive manual curation followed by *in vivo* pathway engineering present themselves as the next steps required downstream of the analysis presented in this project to achieve confirmation of the validity of the nutrient-exchange pathways designed. While a high-throughput approach like the one presented here is well-suited to the transversal of large, diverse design space, there is a trade-off between breadth and depth of scope can be encountered. One practical example of this trade-off is the infeasibility of stoichiometrically balancing all reactions discovered as part of this project, as this task itself was estimated to require over a year of full-time work. The order, in which various experiments were executed, also had to be carefully considered to account for computational capacity and data storage availability. The engineering of a nature-mimicking nutrient exchange system *in vivo* was initially devised to accompany the high-throughput *in silico* approaches. An approach combining both a broad, shallow analysis *in silico* coupled with more granular wet-lab investigation can function in concert to make best use of the benefits of both approaches; unfortunately, work on this part of the project had to be halted due to a sudden onset of a medical condition preventing safe laboratory work.

8.3 Consortium Robustness to Invasion

We have proposed that new- or rare-to-nature nutrient-exchange compounds may have the advantage of being less amenable to breakdown by other members of the ecosystem, hence reducing the dissipation of resources into the environment and improving the robustness of the symbiotic system to competition. While pursuing this project, I have investigated the option of implementing an "invader" into the consortium genome-scale metabolic model in the form of another GEM (for example that of *E. coli*) capable of taking up nutrients produced by the *A. vinelandii*-*Synechocystis* 6803 consortium. This approach was however deemed too limited to yield informative results in the case of any novel nutrient-exchange metabolites. While any nutrients already implemented as part of the invader GEM would likely be amenable to breakdown, any novel compound would simply result in an orphan reaction disconnected from the native metabolism. *In vivo*, however, enzyme promiscuity or reactions which had simply not included in the invader GEM to date are likely to result in a far smaller set of nutrient exchange compounds proving robust to competition than a simple model invader would predict. Instead, I propose that future work could focus on combining feeding assays, in which a candidate nutrient exchange compound is provided in place of a standard source of the nutrient in question to an invader under investigation, with adaptive laboratory evolution experiments (ALE). ALE may be particularly useful to evaluating the performance of nutrient exchange consortium *in situ*, as any eventual downstream application of the symbiotic consortium would be exposed to potential invaders over an extended timeframe, during which competing members of the ecosystem which may be more compatible with the nutrient exchange compound in question can be inadvertently selected for.

8.4 Select Pathways for Downstream *In Vitro* Evaluation

The nutrient exchange pathways identified as part of this project represent starting points for downstream *in vitro* pathway assembly to confirm the validity of the designed nutrient-exchange pathways. A filtered subset of exchange pathway combinations which may be used to guide a downstream experimental project is presented in Table 46. See Supplementary data, Sections 9.1.1 and 9.1.2 for a full graphical representation of all unique exchange pathways in this filtered subset.

First, pathways for which no enzyme with the appropriate catalytic activity had been identified were filtered out by excluding all pathway combinations where any reaction was missing an EC number from the data set.

Next, pathways were filtered based on the feasibility of assembly. The biochemical score, used to determine reaction "trustworthiness" (Koch et al., 2019) as outlined in Section 6.3.1.1, was averaged across the entire pathway. To penalise longer pathways which may prove less practical to assemble *in vitro*, the averaged pathway biochemical score is then divided by the number of reactions in a pathway (n) to calculate the pathway priority score:

$$score_{biochemistry} = score_{biology} \times score_{chemistry}$$

$$score_{pathway\ biochemistry} = \frac{1}{n}\sum_{i=1}^{n}(score_{biology} \times score_{chemistry})$$

$$score_{pathway\ priority} = \frac{score_{pathway\ biochemistry}}{n}$$

The pathway priority score was next averaged across the carbon exchange pathway and the nitrogen exchange pathway for each combination. Pathway combinations falling into the top half of the data set for average pathway priority score are more likely to have the correct activity *in vitro* than the average pathway combination in the dataset.

Finally, the pathway combinations ranking in top 5% maximum biomass flux for both species were identified. For clarity, I have prepared independent subsets representing the top 5% maximum biomass flux for *Synechocystis* and for *Azotobacter*. Only such pathway combinations present in both subsets were carried forward. The final filtered set of pathway combinations contains 76 pathway combinations comprising 52 unique carbon-carrying pathways and 26 unique nitrogen-carrying pathways.

Table 46: Top filtered pathway combinations for downstream *in vitro* evaluation. Carbon and nitrogen exchange pathways are identified using a pathway ID, for a full graphical representation of pathways, see Supplementary data, Sections 9.1.1 and 9.1.2.

Synechocystis **Optimum (gDW/h)**	***Azotobacter*** **Optimum (gDW/h)**	**N Exchange Pathway**	**C Exchange Pathway**	**N Exchange Compound**	**C Exchange Compound**
1.4658	3.783	N117702	C132249	3-cyano-L-alanine	glycogen
0.9692	3.783	N103647	C132249	5-oxo-D-proline	glycogen
0.9692	3.783	N103648	C132249	5-oxo-D-proline	glycogen
0.9692	3.783	N78437	C132249	2-(S)-3-(S)-3-methyl-L-aspartate	glycogen
0.793	3.783	N99387	C132249	3-chloro-L-alanine	glycogen
0.7797	3.783	N103651	C132249	5-oxo-D-proline	glycogen
0.6675	3.783	N103689	C132249	5-oxo-D-proline	glycogen
0.5636	3.783	N103654	C132249	5-oxo-D-proline	glycogen
0.4991	3.783	N96555	C132249	2-aminopropanamide	glycogen
0.3811	3.783	N91117	C132249	cyanate	glycogen
0.3427	3.783	N91057	C132249	cyanate	glycogen
0.6793	2.6728	N117702	C190400	3-cyano-L-alanine	xylitol
0.6407	2.6728	N103647	C190400	5-oxo-D-proline	xylitol
0.5268	2.6728	N103652	C190400	5-oxo-D-proline	xylitol
0.4933	2.6728	N103650	C190400	5-oxo-D-proline	xylitol
0.3849	2.6728	N91112	C190400	cyanate	xylitol
0.3569	2.6728	N91057	C190400	cyanate	xylitol
0.3303	0.9919	N78437	C216651	2-(S)-3-(S)-3-methyl-L-aspartate	1-phosphoalpha-D-galacturonate
0.4111	0.9619	N103652	C45584	5-oxo-D-proline	37-dideoxy-D-threohepto-26-diuolosonate
0.3828	0.9619	N103652	C45662	5-oxo-D-proline	37-dideoxy-D-threohepto-26-diuolosonate
0.1711	0.9515	N103652	C237451	5-oxo-D-proline	2-oxoadipate
0.4518	0.9404	N103694	C173913	5-oxo-D-proline	D-ribono-14-lactone

Synechocystis Optimum (gDW/h)	*Azotobacter* Optimum (gDW/h)	N Exchange Pathway	C Exchange Pathway	N Exchange Compound	C Exchange Compound
0.8538	0.9308	N128624	C105522	4-methyl-L-glutamate-1	4-Methylene-2-oxoglutarate
0.793	0.9308	N128624	C105692	4-methyl-L-glutamate-1	4-Methylene-2-oxoglutarate
0.793	0.9308	N128624	C105590	4-methyl-L-glutamate-1	4-Methylene-2-oxoglutarate
0.5498	0.9212	N103648	C280087	5-oxo-D-proline	3-deoxy-alpha-D-mannooct-2-ulosonate
0.4496	0.9165	N109258	C39498	2-hydroxysuccinamate	D-glycero-D-altrooctulose-18-bisphosphate
0.4939	0.9126	N103650	C209823	5-oxo-D-proline	L-xylono-14-lactone
0.8674	0.9076	N103648	C39462	5-oxo-D-proline	D-glycero-D-altrooctulose-18-bisphosphate
0.2572	0.8998	N109258	C39612	2-hydroxysuccinamate	D-glycero-D-altrooctulose-18-bisphosphate
0.505	0.8903	N91117	C135906	cyanate	3-maleylpyruvate
0.8752	0.8883	N96555	C207651	2-aminopropanamide	2-(R)-2-O-alpha-D-mannosylglycerate
0.3458	0.8652	N78437	C237455	2-(S)-3-(S)-3-methyl-L-aspartate	2-oxoadipate
0.3376	0.8617	N128638	C106024	4-methyl-L-glutamate-1	Cis-4-coumarate
0.7737	0.8575	N117702	C238429	3-cyano-L-alanine	2-oxoadipate
0.3282	0.8575	N103651	C173913	5-oxo-D-proline	D-ribono-14-lactone
0.4531	0.8535	N103692	C39472	5-oxo-D-proline	D-glycero-D-altrooctulose-18-bisphosphate
0.207	0.8314	N96555	C181313	2-aminopropanamide	2-(R)-2-O-alpha-D-mannopyranosyl-12-alpha-D-glucopyranosylglycerate
0.2819	0.8231	N99366	C216639	Glycyl-Glutamine	1-phospho-alpha-D-galacturonate
0.1669	0.8123	N99387	C155050	3-chloro-L-alanine	2-methylpropanal
0.3434	0.8091	N103654	C215645	5-oxo-D-proline	34-dihydroxybenzaldehyde
0.1581	0.7611	N103651	C200531	5-oxo-D-proline	2-keto-5-phospho-3-cispentenoate
0.2225	0.7464	N103652	C113727	5-oxo-D-proline	2-methyleneglutarate
0.6413	0.7434	N103651	C216642	5-oxo-D-proline	1-phosphoalpha-D-galacturonate
0.5847	0.7374	N109258	C216653	2-hydroxysuccinamate	1-phosphoalpha-D-galacturonate
0.7272	0.718	N49737	C39454	2-hydroxy-5-oxoproline	D-glycero-D-altrooctulose-18-bisphosphate
0.1995	0.7122	N78437	C113724	2-(S)-3-(S)-3-methyl-L-aspartate	2-methyleneglutarate

Synechocystis Optimum (gDW/h)	*Azotobacter* Optimum (gDW/h)	N Exchange Pathway	C Exchange Pathway	N Exchange Compound	C Exchange Compound
0.6815	0.7085	N91088	C216653	cyanate	1-phospho-alpha-D-galacturonate
0.3395	0.7075	N103648	C239696	5-oxo-D-proline	2-oxoadipate
0.6404	0.6738	N91117	C177917	cyanate	3-(Z)-2-oxo-4-carboxy-3-hexenedioate
0.2373	0.6656	N103657	C146421	5-oxo-D-proline	2-(R)-2-O-alpha-D-glucopyranosylglycerate
0.2867	0.6569	N103647	C146426	5-oxo-D-proline	2-(R)-2-O-alphaDglucopyranosylglycerate
0.1533	0.6519	N99366	C237455	Glycyl-Glutamine	2-oxoadipate
0.2366	0.6508	N128625	C122489	4-methyl-L-glutamate-1	2-(Z)-4-(Z)-2-hydroxy-5-carboxymuconate-6-semialdehyde
0.2953	0.6451	N96555	C181479	2-aminopropanamide	2-(R)-2-O-alpha-D-mannopyranosyl-12-alpha-D-glucopyranosylglycerate
0.2965	0.6421	N49720	C105505	2-hydroxy-5-oxoproline	4-Methylene-2-oxoglutarate
0.2571	0.6383	N103694	C278608	5-oxo-D-proline	Cis-45-dihydroxycyclohexa-26-diene-12-dicarboxylate
0.2345	0.6206	N103652	C106024	5-oxo-D-proline	Cis-4-coumarate
0.3698	0.5941	N103652	C146461	5-oxo-D-proline	2-(R)-2-O-alpha-D-glucopyranosylglycerate
0.2086	0.5892	N99366	C155695	Glycyl-Glutamine	2-E-2-butenoate
0.7587	0.5801	N99387	C216637	3-chloro-L-alanine	1-phospho-alpha-D-galacturonate
0.258	0.5682	N99366	C173915	Glycyl-Glutamine	D-ribono-14-lactone
0.8503	0.564	N103652	C39612	5-oxo-D-proline	D-glycero-D-altrooctulose-18-bisphosphate
0.5133	0.5635	N103692	C39468	5-oxo-D-proline	D-glycero-D-altrooctulose-18-bisphosphate
0.9061	0.5546	N99387	C108795	3chloroLalanine	L-arabinono-14-lactone
0.9033	0.5431	N103654	C101263	5-oxo-D-proline	(S)-3hydroxybutanoate
0.829	0.5392	N49737	C279252	2-hydroxy-5-oxoproline	Cis-trans-muconate
0.2239	0.5351	N49737	C155695	2-hydroxy-5-oxoproline	2-E-2-butenoate
0.3149	0.5311	N103647	C107013	5-oxo-D-proline	4-hydroxyhexan-3-one
0.2358	0.5229	N103693	C39448	5-oxo-D-proline	D-glycero-D-altrooctulose-18-bisphosphate
0.1555	0.5139	N78437	C215645	2-(S)-3-(S)-3-methyl-L-aspartate	3-4-dihydroxybenzaldehyde

Synechocystis **Optimum (gDW/h)**	***Azotobacter*** **Optimum (gDW/h)**	**N Exchange Pathway**	**C Exchange Pathway**	**N Exchange Compound**	**C Exchange Compound**
0.2101	0.4991	N103650	C180454	5-oxo-D-proline	Aldehyde-D-galacturonate
0.5502	0.4938	N117702	C237574	3-cyano-L-alanine	2-oxoadipate
0.1665	0.4937	N49720	C181313	2-hydroxy-5-oxoproline	2-(R)-2-O-alpha-D-mannopyranosyl-12-alpha-D-glucopyranosylglycerate
0.1817	0.4921	N103689	C155050	5-oxo-D-proline	2-methylpropanal
0.2945	0.4858	N91117	C137485	cyanate	ethene

References

Abdullahi, A. A., Howieson, J., O'Hara, G., Terpolilli, J., Tiwari, R. & Yusuf, A. A. (2020) History of Rhizobia Inoculants Use for Improving Performance of Grain Legumes Based on Experience from Nigeria. *Just Enough Nitrogen.* Springer International Publishing, pp. 101-113.

Ai, L., Muggleton, S. H., Liang, S.-s. & Baldwin, G. S. (2024) Active learning of digenic functions with boolean matrix logic programming. *arXiv: 2408.14487.* https://arxiv.org/abs/2408.14487.

Alleman, A. B., Mus, F. & Peters, J. W. (2021) Metabolic Model of the Nitrogen-Fixing Obligate Aerobe Azotobacter vinelandii Predicts Its Adaptation to Oxygen Concentration and Metal Availability. *mBio.* 12 (6). doi: 10.1128/MBIO.02593-21/SUPPL_FILE/MBIO.02593-21-ST003.XLSX.

Anaconda (2024) *Anaconda Software Distribution*. https://anaconda.com.

Arnold, W., Rump, A., Klipp, W., Priefer, U. B. & Pühler, A. (1988) Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *Journal of Molecular Biology.* 203 (3), 715-738.

Aurich, M. K., Fleming, R. M. T. & Thiele, I. (2016) MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models. *Frontiers in Physiology.* 7.

Bajusz, D., Rácz, A. & Héberger, K. (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics.* 7 (1). doi: 10.1186/s13321-015-0069-3.

Bali, A., Blanco, G., Hill, S. & Kennedy, C. (1992) Excretion of ammonium by a nifL mutant of Azotobacter vinelandii fixing nitrogen. *Applied and Environmental Microbiology.* 58 (5), 1711-1718. doi: 10.1128/aem.58.5.1711-1718.1992.

Barney, B. M., Eberhart, L. J., Ohlert, J. M., Knutson, C. M. & Plunkett, M. H. (2015) Gene Deletions Resulting in Increased Nitrogen Release by Azotobacter vinelandii: Application of a Novel Nitrogen Biosensor. *Applied and Environmental Microbiology.* 81 (13), 4316-4328. doi: 10.1128/AEM.00554-15.

Baysal, C., Burén, S., He, W., Jiang, X., Capell, T., Rubio, L. M. & Christou, P. (2022) Functional expression of the nitrogenase Fe protein in transgenic rice. *Communications Biology.* 5 (1). doi: 10.1038/s42003-022-03921-9.

Beatty, P. H. & Good, A. G. (2011). *Future prospects for cereals that fix nitrogen. Science.* American Association for the Advancement of Science. https://www.science.org/doi/10.1126/science.1209467.

Beber, M. E., Gollub, M. G., Mozaffari, D., Shebek, K. M., Flamholz, I., Avi, Milo, R. & Noor, E. (2022) eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Research.* 50 (D1), D603-D609. doi: 10.1093/nar/gkab1106.

Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø. & Herrgard, M. J. (2007) Quantitative prediction of cellular metabolism with contraint-based models: the COBRA Toolbox. *Nature Protocols.* 2.

Behm, J. E., Geurts, R. & Kiers, E. T. (2014) Parasponia: a novel system for studying mutualism stability. *Trends in Plant Science.* 19 (12), 757-763. doi: 10.1016/j.tplants.2014.08.007.

Benchling (2021). *Benchling [Biology Software]*. https://benchling.com.

Bernard, T., Bridge, A., Morgat, A., Moretti, S., Xenarios, I. & Pagni, M. (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. *Briefings in Bioinformatics.* 15 (1), 123-135. http://bib.oxfordjournals.org/content/15/1/123.abstractN2.

Bernhard, A. (2010) The Nitrogen Cycle: Processes, Players, and Human Impact. *Nature Education Knowledge*. 3 (10), 25.
Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K. & Wiswedel, B. (2008) KNIME: The Konstanz Information Miner. *Studies in Classification, Data Analysis, and Knowledge Organization.* Springer Berlin Heidelberg, pp. 319-326.
Bertrani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *Journal of Bacteriology.* 62 (3), 293-300. doi: 10.1128/JB.62.3.293-300.1951.
Boddey, R. M. & Dobereiner, J. (1995) Nitrogen fixation associated with grasses and cereals: Recent progress and perspectives for the future. *Fertilizer Research.* 42 (1-3), 241-250. doi: 10.1007/bf00750518.
Boyd, E. S. & Peters, J. W. (2013) New insights into the evolutionary history of biological nitrogen fixation. *Frontiers in Microbiology.* 4. doi: 10.3389/fmicb.2013.00201.
Bryksin, A. V. & Matsumura, I. (2010) Overlap extension PCR cloning: A simple and reliable way to create recombinant plasmids. *BioTechniques.* 48 (6), 463-465. doi: 10.2144/000113418/ASSET/IMAGES/LARGE/FIGURE2.JPEG.
Bulen, W. A. & Lecomte, J. R. (1966) The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proceedings of the National Academy of Sciences.* 56 (3), 979-986. doi: 10.1073/pnas.56.3.979.
Burén, S., Jiménez-Vicente, E., Echavarri-Erasun, C. & Rubio, L. M. (2020) Biosynthesis of Nitrogenase Cofactors. *Chemical Reviews.* 120 (12).
Burén, S., Young, E. M., Sweeny, E. A., Lopez-Torrejón, G., Veldhuizen, M., Voigt, C. A. & Rubio, L. M. (2017) Formation of Nitrogenase NifDK Tetramers in the Mitochondria of Saccharomyces cerevisiae. *ACS Synthetic Biology.* 6 (6), 1043-1055. doi: 10.1021/acssynbio.6b00371.
Carbonell, P., Parutto, P., Baudier, C., Junot, C. & Faulon, J.-L. (2013) Retropath: Automated Pipeline for Embedded Metabolic Circuits. *ACS Synthetic Biology.* 3 (8).
Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S., Subhraveti, P. & Karp, P. D. (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research.* 46 (D1), D633-D639. doi: 10.1093/nar/gkx935.
Castillo, T., López, I., Flores, C., Segura, D., García, A., Galindo, E. & Peña, C. (2018) Oxygen uptake rate in alginate producer (algU+) and nonproducer (algU−) strains of Azotobacter vinelandii under nitrogen-fixation conditions. *Journal of Applied Microbiology.* 125 (1), 181-189. doi: https://doi.org/10.1111/jam.13760.
Ceballos-Aguirre, N., Cuellar, J. A., Restrepo, G. M. & Sánchez, Ó. J. (2023) Effect of the Application of Gluconacetobacter diazotrophicus and Its Interaction with Nitrogen and Phosphorus Fertilization on Carrot Yield in the Field. *International Journal of Agronomy.* 2023, 1-9. doi: 10.1155/2023/6899532.
Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D. & Schomburg, D. (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research.* 49 (D1), D498-D508. doi: 10.1093/nar/gkaa1025.
Charpentier, M. & Oldroyd, G. (2010). *How close are we to nitrogen-fixing cereals? Current Opinion in Plant Biology.*
Christiansen-Weniger, C. (1992) N2-fixation by ammonium-excreting Azospirillum brasilense in auxine-induced root tumours of wheat (Triticum aestivum L.). *Biology and Fertility of Soils.* 13 (3), 165-172. doi: 10.1007/bf00336274.

Christiansen-Weniger, C. & Vanderleyden, J. (1994) Ammonium-excreting Azospirillum sp. become intracellularly established in maize (Zea mays) para-nodules. *Biology and Fertility of Soils.* 17 (1), 1-8. doi: 10.1007/bf00418663.
Cocking, E. C., Stone, P. J. & Davey, M. R. (2006) Intracellular colonization of roots of Arabidopsis and crop plants by Gluconacetobacter diazotrophicus. *In Vitro Cellular & Developmental Biology - Plant.* 42 (1), 74-82. doi: 10.1079/ivp2005716.
Curatti, L., Hernandez, J. A., Igarashi, R. Y., Soboh, B., Zhao, D. & Rubio, L. M. (2007) *In vitro* synthesis of the iron–molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proceedings of the National Academy of Sciences.* 104 (45), 17626-17631. doi: 10.1073/pnas.0703050104.
Curatti, L. & Rubio, L. M. (2014). *Challenges to develop nitrogen-fixing cereals by direct nif-gene transfer. Plant Science.* Elsevier Ireland Ltd.
Delépine, B., Duigou, T., Carbonell, P. & Faulon, J. L. (2018) RetroPath2.0: A retrosynthesis workflow for metabolic engineers. *Metabolic Engineering.* 45, 158-170. doi: 10.1016/j.ymben.2017.12.002.
Dixon, R., Cheng, Q., Shen, G.-F., Day, A. & Dowson-Day, M. (1997) *Nif* gene transfer and expression in chlorplasts: Prospects and problems. *Plant and Soil.* 194 (1/2), 193-203. doi: 10.1023/a:1004296703638.
Dixon, R. & Kahn, D. (2004). *Genetic regulation of biological nitrogen fixation. Nature Reviews Microbiology.* Nature Publishing Group. https://www.nature.com/articles/nrmicro954.
Dixon, R. A. & Postgate, J. R. (1972) Genetic Transfer of Nitrogen Fixation from Klebsiella pneumoniae to Escherichia coli. *Nature.* 237 (5350), 102-103. doi: 10.1038/237102a0.
Dos Santos, P. C. (2011) Molecular biology and genetic engineering in nitrogen fixation. *Methods in Molecular Biology.* 766, 81-92. doi: 10.1007/978-1-61779-194-9_6.
Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C. & Dixon, R. (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics.* 13 (1), 162-162. doi: 10.1186/1471-2164-13-162.
Du, W., Liang, F., Duan, Y., Tan, X. & Lu, X. (2013) Exploring the photosynthetic production capacity of sucrose by cyanobacteria. *Metabolic Engineering.* 19, 17-25. doi: 10.1016/j.ymben.2013.05.001.
Ducat, D. C., Avelar-Rivas, J. A., Way, J. C. & Silvera, P. A. (2012) Rerouting carbon flux to enhance photosynthetic productivity. *Applied and Environmental Microbiology.* 78 (8), 2660-2668. doi: 10.1128/AEM.07901-11.
Duigou, T., Du Lac, M., Carbonell, P. & Faulon, J. L. (2019) RetroRules: a database of reaction rules for engineering biology. *Nucleic Acids Research.* 47 (D1), D1229-D1235. doi: 10.1093/NAR/GKY940.
Eady, R. R. (1996) Structure-Function Relationships of Alternative Nitrogenases. *Chemical Reviews.* 96 (7).
Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology.* 7 (1), 74-74. doi: 10.1186/1752-0509-7-74.
Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. (2008) How a century of ammonia synthesis changed the world. *Nature Geoscience.* 1 (10), 636-639. doi: 10.1038/ngeo325.
Eseverri, Á., López-Torrejón, G., Jiang, X., Burén, S., Rubio, L. M. & Caro, E. (2020) Use of synthetic biology tools to optimize the production of active nitrogenase Fe protein in chloroplasts of tobacco leaf cells. *Plant Biotechnology Journal.* 18 (9), 1882-1896. doi: 10.1111/pbi.13347.
Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C.,

Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H. & D'Eustachio, P. (2018) The Reactome Pathway Knowledgebase. *Nucleic Acids Research.* 46 (D1), D649-D655. doi: 10.1093/nar/gkx1132.
Fang, X., Lloyd, C. J. & Palsson, B. O. (2020) Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology.* 18 (12), 731-743. doi: 10.1038/s41579-020-00440-4.
Feist, A. M. & Palsson, B. O. (2010) The biomass objective function. *Current Opinion in Microbiology.* 13 (3), 344-349. doi: 10.1016/j.mib.2010.03.003.
Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. (2012) eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Research.* 40 (D1), D770-D775. doi: 10.1093/nar/gkr874.
Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B. & Lercher, M. J. (2017) Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLOS Computational Biology.* 13 (4), e1005494. doi: 10.1371/journal.pcbi.1005494.
Fuentes-Ramirez, L. E., Jimenez-Salgado, T., Abarca-Ocampo, I. R. & Caballero-Mellado, J. (1993) Acetobacter diazotrophicus, an indoleacetic acid producing bacterium isolated from sugarcane cultivars of Mexico. *Plant and Soil.* 154 (2), 145-150. doi: 10.1007/bf00012519.
Gallon, J. R. (1981) The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. *Trends in Biochemical Sciences.* 6, 12-23.
Galloway, J. N., Schlesinger, W. H., Levy, H., Michaels, A. & Schnoor, J. L. (1995) Nitrogen fixation: Anthropogenic enhancement-environmental response. *Global Biogeochemical Cycles.* 9 (2), 235-252. doi: 10.1029/95GB00158.
Ganter, M., Bernard, T., Moretti, S., Stelling, J. & Pagni, M. (2013) MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics.* 29 (6), 815-816. doi: 10.1093/bioinformatics/btt036.
Gricourt, G., Meyer, P., Duigou, T. & Faulon, J.-L. (2024) Artificial Intelligence Methods and Models for Retro-Biosynthesis: A Scoping Review. *ACS Synthetic Biology.* 13 (8), 2276-2294. doi: 10.1021/acssynbio.4c00091.
Haber, F. (1920) *The synthesis of ammonia from its elements*.
Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. & Oliphant, T. E. (2020) Array programming with NumPy. *Nature.* 585 (7825), 357-362. doi: 10.1038/s41586-020-2649-2.
Healy, J. & Mcinnes, L. (2024) Uniform manifold approximation and projection. *Nature Reviews Methods Primers.* 4 (1). doi: 10.1038/s43586-024-00363-x.
Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D. C., Valcarcel, L. V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H. M., Ma, D., Sun, Y., Wang, L., Yurkovich, J. T., Oliveira, M. A. P., Vuong, P. T., El Assal, L. P., Kuperstein, I., Zinovyev, A., Hinton, H. S., Bryant, W. A., Aragón Artacho, F. J., Planes, F. J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M. A., Maranas, C. D., Lewis, N. E., Sauter, T., Palsson, B. Ø., Thiele, I. & Fleming, R. M. T. (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols.* 14 (3), 639-702. doi: 10.1038/s41596-018-0098-2.
Henderson, J. W., Jr. & Brooks, A. (2010) *Improved Amino Acid Methods using Agilent ZORBAX Eclipse Plus C18 Columns for a Variety of Agilent LC Instrumentation and Separation Goals*. Wilmington, USA,

Henderson, J. W., Ricker, R. D., Bidlingmeyer, B. A. & Woodward, C. (2000) *Rapid, Accurate, Sensitive, and Reproducible HPLC Analysis of Amino Acids Amino Acid Analysis Using Zorbax Eclipse-AAA Columns and the Agilent 1100 HPLC*. www.agilent.com/chem.
Hill, D. J. (1975) The pattern of development of Anabaena in the Azolla-Anabaena symbiosis. *Planta.* 122 (2), 179-184. doi: 10.1007/bf00388657.
Hill, D. J. (1977) The Role of Anabaena in the Azolla-Anabaena Symbiosis. *New Phytologist.* 78, 611-616.
Huang, H.-H., Camsund, D., Lindblad, P. & Heidorn, T. (2022) Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Research.* 38 (8), 2577-2593. doi: 10.1093/nar/gkq164.
Hunter, J. D. (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering.* 9 (3).
InChI Trust (2024) *Welcome to InChI.* https://www.inchi-trust.org/.
Iswaran, V. & Sundara Rao, W. V. B. (1960) The Effect of Trace Elements on Nitrogen Fixation by *Azotobacter*. *Proceedings of the Indian Academy of Sciences.*
Jain, A. & Srivastava, P. (2013) Broad host range plasmids. *FEMS Microbiology Letters.* 348 (2), 87-96. doi: 10.1111/1574-6968.12241.
Jiang, X., Payá-Tormo, L., Coroian, D., García-Rubio, I., Castellanos-Rueda, R., Eseverri, Á., López-Torrejón, G., Burén, S. & Rubio, L. M. (2021) Exploiting genetic diversity and gene synthesis to identify superior nitrogenase NifH protein variants to engineer N2-fixation in plants. *Communications Biology.* 4 (1). doi: 10.1038/s42003-020-01536-6.
Joshi, C. J., Peebles, C. A. M. & Prasad, A. (2017) Modeling and analysis of flux distribution and bioproduct formation in Synechocystis sp. PCC 6803 using a new genome-scale metabolic reconstruction. *Algal Research.* 27, 295-310. doi: 10.1016/J.ALGAL.2017.09.013.
Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA research : an international journal for rapid publication of reports on genes and genomes.* 3 (3), 109-136. doi: 10.1093/DNARES/3.3.109.
Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. (1995) Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA research : an international journal for rapid publication of reports on genes and genomes.* 2 (4), 153-166. doi: 10.1093/DNARES/2.4.153.
Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. (2003) Host sanctions and the legume–rhizobium mutualism. *Nature.* 425 (6953), 78-81. doi: 10.1038/nature01931.
Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J. & Bolton, E. E. (2023) PubChem 2023 update. *Nucleic Acids Research.* 51 (D1), D1373-D1380. doi: 10.1093/nar/gkac956.
King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O. & Lewis, N. E. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research.* 44 (D1), D515-D522. doi: 10.1093/nar/gkv1049.
Koch, M. (2019) *Github: RetroPathRL/document_all_options.md.* https://github.com/brsynth/RetroPathRL/blob/master/document_all_options.md.
Koch, M., Duigou, T. & Faulon, J.-L. (2019) Reinforcement Learning for Bioretrosynthesis. *ACS Synthetic Biology.* 9 (1), 157-168. doi: 10.1021/acssynbio.9b00447.

Koning, N. B. J., Van Ittersum, M. K., Becx, G. A., Van Boekel, M. A. J. S., Brandenburg, W. A., Van Den Broek, J. A., Goudriaan, J., Van Hofwegen, G., Jongeneel, R. A., Schiere, J. B. & Smies, M. (2008) Long-term global availability of food: continued abundance or new scarcity? *NJAS: Wageningen Journal of Life Sciences.* 55 (3), 229-292. doi: 10.1016/s1573-5214(08)80001-2.
Kumar, K., Mella-Herrera, R. A. & Golden, J. W. (2010) Cyanobacterial Heterocysts. *Cold Spring Harbor Perspectives in Biology.* 2 (4), a000315-a000315. doi: 10.1101/cshperspect.a000315.
Laudrum, G. & others (2024) *The RDKit Book.* https://www.rdkit.org/docs/RDKit_Book.html#molecular-sanitization.
Leach, A. R., Bradshaw, J., Green, D. V. S., Hann, M. M. & Delany, J. J. (1999) Implementation of a System for Reagent Selection and Library Enumeration, Profiling, and Design. *Journal of Chemical Information and Computer Sciences.* 39 (6), 1161-1172.
Long, S. R. (1996) Rhizobium symbiosis: nod factors in perspective. *Plant Cell.* 8 (10).
López-Torrejón, G., Jiménez-Vicente, E., Buesa, J. M., Hernandez, J. A., Verma, H. K. & Rubio, L. M. (2016) Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nature Communications.* 7 (1), 1-6. doi: 10.1038/ncomms11426.
Lupwayi, N. Z., Clayton, G. W., Hanson, K. G., Rice, W. A. & Biederbeck, V. O. (2004) Endophytic rhizobia in barley, wheat and canola roots. *Canadian Journal of Plant Science.*
Lynch, M. & Marinov, G. K. (2015) The bioenergetic costs of a gene. *PNAS.* 112 (51), 15690-15695.
Ma, F., Jazmin, L. J., Young, J. D. & Allen, D. K. (2014) Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences.* 111 (47), 16967-16972.
Mackellar, D., Lieber, L., Norman, J. S., Bolger, A., Tobin, C., Murray, J. W., Oksaksin, M., Chang, R. L., Ford, T. J., Nguyen, P. Q., Woodward, J., Permingeat, H. R., Joshi, N. S., Silver, P. A., Usadel, B., Rutherford, A. W., Friesen, M. L. & Prell, J. (2016) Streptomyces thermoautotrophicus does not fix nitrogen. *Scientific Reports.* 6 (1), 20086. doi: 10.1038/srep20086.
Malatinszky, D., Steuer, R. & Jones, P. R. (2017) A comprehensively curated genome-scale two-cell model for the heterocystous cyanobacterium Anabaena sp. PCC 7120. *Plant Physiology.* 173 (1), 509-523. doi: 10.1104/pp.16.01487.
Martin Del Campo, J. S., Rigsbee, J., Bueno Batista, M., Mus, F., Rubio, L. M., Einsle, O., Peters, J. W., Dixon, R., Dean, D. R. & Dos Santos, P. C. (2023) Overview of physiological, biochemical, and regulatory aspects of nitrogen fixation in Azotobacter vinelandii. *Critical Reviews in Biochemistry and Molecular Biology.* 57 (5-6), 492-538. doi: 10.1080/10409238.2023.2181309.
Matsudaira, A., Hoshino, Y., Uesaka, K., Takatani, N., Omata, T. & Usuda, Y. (2020) Production of glutamate and stereospecific flavors, (S)-linalool and (+)-valencene, by Synechocystis sp. PCC6803. *Journal of Bioscience and Bioengineering.* 130 (5), 464-470. doi: 10.1016/j.jbiosc.2020.06.013.
Mcinnes, L., Healy, J. & Astels, S. (2017) hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software.* 2 (11), 205. doi: 10.21105/joss.00205.
McInnes, L., Healy, J. & Melville, J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426.* https://arxiv.org/abs/1802.03426.
Mcintosh, C. L., Germer, F., Schulz, R., Appel, J. & Jones, A. K. (2011) The [NiFe]-Hydrogenase of the Cyanobacterium Synechocystis sp. PCC 6803 Works Bidirectionally with a Bias to H₂ Production. *Journal of the American Chemical Society.* 133 (29), 11308-11319. doi: 10.1021/ja203376y.

McKinney, W. (2010) *Data structures for statistical computing in python*, *Proceedings of the 9th Python in Science Conference*.
Meeks, J. C. & Elhai, J. (2002) Regulation of Cellular Differentiation in Filamentous Cyanobacteria in Free-Living and Plant-Associated Symbiotic Growth States. *Microbiology and Molecular Biology Reviews.* 66 (1), 94-121. doi: 10.1128/mmbr.66.1.94-121.2002.
Menon, A. L. & Robson, R. L. (1994) In vivo and in vitro nickel-dependent processing of the [NiFe] hydrogenase in Azotobacter vinelandii. *Journal of Bacteriology.* 176 (2).
Merrick, M. & Dixon, R. (1984) Why don't plants fix nitrogen? *Trends in Biotechnology.* 2 (6), 162-166.
Mishra, P., Roy, B., Prasad, R. & Das, H. K. (1991) Amino acid transport in Azotobacter vinelandii: implications of nonavailability of amino acid auxotrophs. *FEMS Microbiology Letters.* 79, 41-44. doi: 10.1111/j.1574-6968.1991.tb04501.x.
Monge-Acuña, A. A. & Fornaguera-Trías, J. (2009) A high performance liquid chromatography method with electrochemical detection of gamma-aminobutyric acid, glutamate and glutamine in rat brain homogenates. *Journal of Neuroscience Methods.* 183 (2), 176-181. doi: 10.1016/J.JNEUMETH.2009.06.042.
Monk, J., Nogales, J. & Palsson, B. O. (2014) Optimizing genome-scale network reconstructions. *Nature Biotechnology.* 32 (5), 447-452. doi: 10.1038/nbt.2870.
Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A. M. & Palsson, B. O. (2017) iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology.* 35 (10), 904-908. doi: 10.1038/nbt.3956.
Moretti, S., Martin, O., Van Du, T., Bridge, A., Morgat, A. & Pagni, M. (2016a) MetaNetX/MNXref - reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research.* 44 (D1), D523-D526. http://nar.oxfordjournals.org/content/44/D1/D523.abstract.
Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A. & Pagni, M. (2016b) MetaNetX/MNXref - Reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research.* 44 (D1), D523-D526. doi: 10.1093/nar/gkv1117.
Moretti, S., Tran, V. D. T., Mehl, F., Ibberson, M. & Pagni, M. (2021) MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Research.* 49 (D1), D570-D574. doi: 10.1093/nar/gkaa992.
Morgat, A., Lombardot, T., Axelsen, K. B., Aimo, L., Niknejad, A., Hyka-Nouspikel, N., Coudert, E., Pozzato, M., Pagni, M., Moretti, S., Rosanoff, S., Onwubiko, J., Bougueleret, L., Xenarios, I., Redaschi, N. & Bridge, A. (2017) Updates in Rhea – an expert curated resource of biochemical reactions. *Nucleic Acids Research.* 45 (D1), D415-D418. doi: 10.1093/nar/gkw990.
Mus, F., Crook, M. B., Garcia, K., Costas, A. G., Geddes, B. A., Kouri, E. D., Paramasivan, P., Ryu, M. H., Oldroyd, G. E. D., Poole, P. S., Udvardi, M. K., Voigt, C. A., Ané, J. M. & Peters, J. W. (2016). *Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. Applied and Environmental Microbiology.* American Society for Microbiology. http://aem.asm.org/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4907175/pdf/zam3698.pdf.
Myers, J. A., Curtis, B. S. & Curtis, W. R. (2013) Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophysics.* 6 (1), 4. doi: 10.1186/2046-1682-6-4.

Nagargade, M., Tyagi, V. & Singh, M. K. (2018) Plant Growth-Promoting Rhizobacteria: A Biological Approach Toward the Production of Sustainable Agriculture. *Role of Rhizospheric Microbes in Soil.* Springer Singapore, pp. 205-223.
Nam, H., Lewis, N. E., Lerman, J. A., Lee, D.-H., Chang, R. L., Kim, D. & Palsson, B. O. (2012) Network Context and Selection in the Evolution to Enzyme Specificity. *Science.* 337 (6098), 1101-1104. doi: 10.1126/science.1216861.
Noor, E., Bar-Even, A., Flamholz, A., Lubling, Y., Davidi, D. & Milo, R. (2012) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics.* 28 (15), 2037-2044. doi: 10.1093/bioinformatics/bts317.
Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W. & Milo, R. (2014) Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism. *PLOS Computational Biology.* 10 (2), e1003483. doi: 10.1371/journal.pcbi.1003483.
Noor, E., Haraldsdóttir, H. S., Milo, R. & Fleming, R. M. T. (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. *PLOS Computational Biology.* 9 (7), e1003098. doi: 10.1371/journal.pcbi.1003098.
O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T. & Hutchison, G. R. (2011) Open Babel: An open chemical toolbox. *Journal of Cheminformatics.* 3 (1), 33. doi: 10.1186/1758-2946-3-33.
Oldroyd, G. E. D. & Downie, J. A. (2008) Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annual Review of Plant Biology.* 59, 519-546.
Oldroyd, G. E. D., Murray, J. D., Poole, P. S. & Downie, J. A. (2011) The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annual Review of Genetics.* 45.
Op Den Camp, R. H. M., Polone, E., Fedorova, E., Roelofsen, W., Squartini, A., Op Den Camp, H. J. M., Bisseling, T. & Geurts, R. (2012) Nonlegume *Parasponia andersonii* Deploys a Broad Rhizobium Host Range Strategy Resulting in Largely Variable Symbiotic Effectiveness. *Molecular Plant-Microbe Interactions®.* 25 (7), 954-963. doi: 10.1094/mpmi-11-11-0304.
Orosz, A., Boros, I. & Venetianer, P. (1991) Analysis of the complex transcription termination region of the Escherichia coli rrn B gene. *European Journal of Biochemistry.* 201 (3), 653-659. doi: 10.1111/j.1432-1033.1991.tb16326.x.
Orth, J. D., Fleming, R. M. T. & Palsson, B. Ø. (2010a) Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal Plus.* 4 (1). doi: 10.1128/ECOSALPLUS.10.2.1/ASSET/AF2F042E-12AF-48A1-967D-9627A8D81014/ASSETS/GRAPHIC/10.2.1_FIG_022.GIF.
Orth, J. D., Thiele, I. & Palsson, B. Ø. (2010b) What is flux balance analysis? *Nature Biotechnology.* 28 (3), 245-248. doi: 10.1038/nbt.1614.
Page, W. J. & Von Tigerstrom, M. (1979) Optimal conditions for transformation of Azotobacter vinelandii. *Journal of Bacteriology.* 139 (3), 1058-1061. doi: 10.1128/JB.139.3.1058-1061.1979/FORMAT/PDF.
Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M. & Vrgoč, D. (2016) *Foundations of JSON schema*, *Proceedings of the 25th International Conference on World Wide Web*.
Pinzon, W., Vega, H., Gonzalez, J. & Pinzon, A. (2019) Mathematical Framework Behind the Reconstruction and Analysis of Genome Scale Metabolic Models. *Archives of Computational Methods in Engineering.* 26 (5), 1593-1606. doi: 10.1007/s11831-018-9290-3.
Porter, S. S., Dupin, S. E., Denison, R. F., Kiers, E. T. & Sachs, J. L. (2024) Host-imposed control mechanisms in legume–rhizobia symbiosis. *Nature Microbiology.* 9 (8), 1929-1939. doi: 10.1038/s41564-024-01762-2.
Price, N. D., Reed, J. L. & Palsson, B. Ø. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology.* 2 (11), 886-897. doi: 10.1038/nrmicro1023.
RDKit: Open-source cheminformatics. https://www.rdkit.org.

Reddy, P. M., Ladha, J. K., So, R. B., Hernandez, R. J., Ramos, M. C., Angeles, O. R., Dazzo, F. B. & De Bruijn, F. J. (1997) Rhizobial communication with rice roots: Induction of phenotypic changes,mode of invasion and extent of colonization. *Plant and Soil.* 194 (1/2), 81-98. doi: 10.1023/a:1004243915997.
Reed, S. (2011) Functional biology of heterotrophic nitrogen-fixing bacteria. *Annual Review of Ecology, Evolution, and Systematics.* 42, 489-512. doi: 10.1146/annurev-ecolsys-102710-145034.
Reinhold-Hurek, B. & Hurek, T. (1998) Interactions of Gramineous Plants with Azoarcus spp. and Other Diazotrophs: Identification, Localization, and Perspectives to Study their Function. *Critical Reviews in Plant Sciences.* 17 (1), 29-54. doi: 10.1080/07352689891304186.
Rikkinen, J., Oksanen, I. & Lohtander, K. (2002) Lichen Guilds Share Related Cyanobacterial Symbionts. *Science.* 297 (5580), 357.
Rippka, R. (1972) Photoheterotrophy and chemoheterotrophy among unicellular blue-green algae. *Archiv für Mikrobiologie.* 87 (1), 93-98. doi: 10.1007/BF00424781.
Rippka, R. & Stanier, R. Y. (1978) The Effects of Anaerobiosis on Nitrogenase Synthesis and Heterocyst Development by Nostocacean Cyanobacteria. *Microbiology.* 105 (1).
Rogers, D. & Hahn, M. (2010) Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling.* 50 (5).
Rubio, L. M. & Ludden, P. W. (2008). *Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Annual Review of Microbiology.* https://www.annualreviews.org/doi/pdf/10.1146/annurev.micro.62.081307.162737.
Segler, M. H. S., Preuss, M. & Waller, M. P. (2018) Planning chemical syntheses with deep neural networks and symbolic AI. *Nature 2018 555:7698.* 555 (7698), 604-610. doi: 10.1038/nature25978.
Setubal, J. C., Dos Santos, P., Goldman, B. S., Ertesvåg, H., Espin, G., Rubio, L. M., Valla, S., Almeida, N. F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z., Godsy, E., Goodner, B., Hellner-Burris, K., Hernandez, J. A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T. J., Latreille, P., Ligon, L. S., Lu, J., Mærk, M., Miller, N. M., Norton, S., O'Carroll, I. P., Paulsen, I., Raulfs, E. C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S. L., Studholme, D. J., Sun, J., Viana, C. J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D. R., Dixon, R. & Wood, D. (2009) Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *Journal of Bacteriology.* 191 (14), 4534-4545. doi: 10.1128/JB.00504-09.
Shafaei, A., Halim, N. H. A., Zakaria, N. & Ismail, Z. (2017) Analysis of Free Amino Acids in Different Extracts of Orthosiphon stamineus Leaves by High-Performance Liquid Chromatography Combined with Solid-Phase Extraction. *Pharmacognosy magazine.* 13 (Suppl 3), S385-S391. doi: 10.4103/0973-1296.216337.
Shastri, A. A. & Morgan, J. A. (2005) Flux Balance Analysis of Photoautotrophic Metabolism. *Biotechnology Progress.* 21 (6), 1617-1626. doi: 10.1021/bp050246d.
Singh, D., Ghosh, P., Kumar, J. & Kumar, A. (2019) Plant Growth-Promoting Rhizobacteria (PGPRs): Functions and Benefits. *Microbial Interventions in Agriculture and Environment.* Springer Singapore, pp. 205-227.
Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., Woodruff, L. B. A., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Densmore, D. & Voigt, C. A. (2014) Functional optimization of gene clusters by combinatorial design and assembly. *Nature Biotechnology.* 32, 1241-1249.
Smercina, D. N., Evans, S. E., Friesen, M. L. & Tiemann, L. K. (2019). *To fix or not to fix: Controls on free-living nitrogen fixation in the rhizosphere. Applied and Environmental Microbiology.* American Society for Microbiology. http://aem.asm.org/.

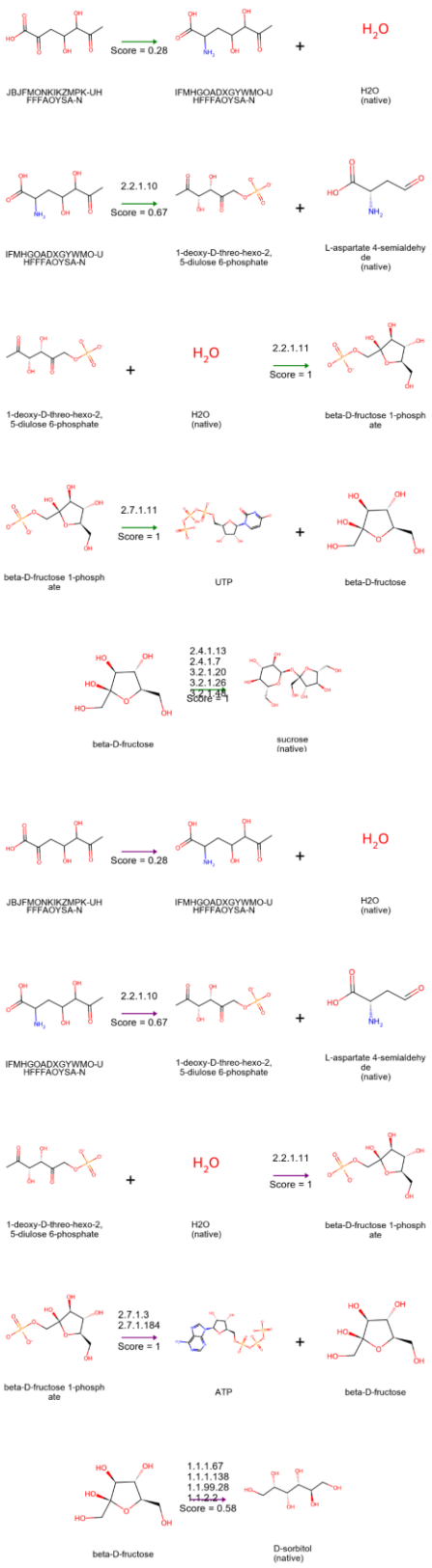
Smith, M. J. & Francis, M. B. (2016) A Designed A. vinelandii – S. elongatus Coculture for Chemical Photoproduction from Air, Water, Phosphate, and Trace Metals. *ACS Synthetic Biology.* 5 (9), 955-961. doi: 10.1021/acssynbio.6b00107.
Stam, H., Van Verseveld, H. W., De Vries, W. & Stouthamer, A. H. (1984) Hydrogen oxidation and efficiency of nitrogen fixation in succinate-limited chemostat cultures ofRhizobium ORS 571. *Archives of Microbiology.* 139 (1), 53-60. doi: 10.1007/bf00692712.
Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T. & Baldwin, G. S. (2015) BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS Synthetic Biology.* 4 (7), 781-787. doi: 10.1021/sb500356d.
Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., Palenik, B., Peterson, T. C., Haerizadeh, F., Golden, S. S. & Golden, J. W. (2022) Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. *Nucleic Acids Research.* 42 (17). doi: 10.1093/nar/gku673.
Tchan, Y. T., Zeman, A. M. M. & Kennedy, I. R. (1991) Nitrogen fixation inpara-nodules of wheat roots by introduced free-living diazotrophs. *Plant and Soil.* 137 (1), 43-47. doi: 10.1007/bf02187430.
Tec-Campos, D., Zuñiga, C., Passi, A., Del Toro, J., Tibocha-Bonilla, J. D., Zepeda, A., Betenbaugh, M. J. & Zengler, K. (2020) Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph *Azotobacter vinelandii* DJ. *Metabolic Engineering Communications.* 11 (e00132).
Tel-Or, E. & Stewart, W. D. P. (1977) Photosynthetic Components and Activities of Nitrogen-Fixing Isolated Heterocysts of *Anabaena cylindrica*. *Proceedings of the Royal Society of London. Series B, Biological Sciences.* 198 (1130).
The UniProt, C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research.* 47 (D1), D506-D515. doi: 10.1093/nar/gky1049.
Thermo Fisher Scientific Inc. (2006-2024) *CloneJET PCR Cloning Kit.* https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-molecular-cloning/clonejet-pcr-cloning-kit.html?ef_id=Cj0KCQiArby5BhCDARIsAIJvjISa9jXjpWaqb1XrZ7Tvs0jK8afrOHkeNaFBhvKAlzUDYC3YYOzBd6saAkdcEALw_wcB:G:s&s_kwcid=AL!3652!3!593223749655!e!!g!!clonejet!10552817726!112107383962&cid=bid_mol_clo_r01_co_cp1358_pjt0000_bid00000_0se_gaw_bt_pur_con&gad_source=1&gclid=Cj0KCQiArby5BhCDARIsAIJvjISa9jXjpWaqb1XrZ7Tvs0jK8afrOHkeNaFBhvKAlzUDYC3YYOzBd6saAkdcEALw_wcB.
Thiele, I. & Palsson, B. Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols.* 5 (1), 93-121. doi: 10.1038/nprot.2009.203.
Tirichine, L., James, E. K., Sandal, N. & Stougaard, J. (2007) Spontaneous Root-Nodule Formation in the Model Legume Lotus japonicus: A Novel Class of Mutants Nodulates in the Absence of Rhizobia. *IS-MPMI.* 19 (4).
Turner, T. R., James, E. K. & Poole, P. S. (2013) The plant microbiome. *Genome Biology.* 14 (6), 1-10. doi: doi:10.1186/gb-2013-14-6-209.
van Heeswijk, W. C., Westerhoff, H. V. & Boogerd, F. C. (2013) Nitrogen Assimilation in Escherichia coli: Putting Molecular Data into a Systems Perspective. *Microbiology and Molecular Biology Reviews.* 77 (4), 628-695. doi: 10.1128/mmbr.00025-13.
van Rhijn, P., Goldberg, R. B. & Hirsch, A. M. (1998) Lotus corniculatus Nodulation Specificity Is Changed by the Presence of a Soybean Lectin Gene. *The Plant Cell.* 10 (8).
Varma, A. & Palsson, B. O. (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology.* 12 (10), 994-998. doi: 10.1038/nbt1094-994.
Vermass, W. F. J., Rutherford, A. W. & Hansson, Ö. (1988) Site-directed mutagenesis in photosystem II of the cyanobacterium *Synechocystis sp*. PCC 6803: Donor D is a

tyrosine residue in the D2 protein. *Proceedings of the National Academy of Sciences.* 85 (22), 8477-8481. doi: 10.1073/pnas.85.22.8477.
Vessey, J. K. (2003) *Plant and Soil.* 255 (2), 571-586. doi: 10.1023/a:1026037216893.
Walker, L., Lagunas, B. & Gifford, M. L. (2020) Determinants of Host Range Specificity in Legume-Rhizobia Symbiosis. *Frontiers in Microbiology.* 11. doi: 10.3389/fmicb.2020.585749.
Walsby, A. E. (1985) The permeability of heterocysts to the gases nitrogen and oxygen. *Proceedings of the Royal Society of London B.*
Wang, L., Xu, R., Hu, B., Li, W., Sun, Y., Tu, Y. & Zeng, X. (2010) Analysis of free amino acids in Chinese teas and flower of tea plant by high performance liquid chromatography combined with solid-phase extraction. *Food Chemistry.* 123 (4), 1259-1266. doi: 10.1016/J.FOODCHEM.2010.05.063.
Waskom, M. (2021) seaborn: statistical data visualization. *Journal of Open Source Software.* 6 (60), 3021. doi: 10.21105/joss.03021.
Willet, P., Barnard, J. M. & Downs, G. M. (1998) Chemical Similarity Searching. *Journal of Chemical Information and Computer Sciences.* 38, 983-996.
Woodard, H. J. & Bly, A. (2000) Maize growth and yield responses to seed-inoculated N_2-fixing bacteria under Dryland production conditions. *Journal of Plant Nutrition.* 23 (1), 55-65. doi: 10.1080/01904160009381997.
Wu, C., Herold, R. A., Knoshaug, E. P., Wang, B., Xiong, W. & Laurens, L. M. L. (2019) Fluxomic Analysis Reveals Central Carbon Metabolism Adaptation for Diazotroph Azotobacter vinelandii Ammonium Excretion. *Scientific Reports.* 9 (1). doi: 10.1038/s41598-019-49717-6.
Yamamoto, C., Toyoshima, M., Kitamura, S., Ueno, Y., Akimoto, S., Toya, Y. & Shimizu, H. (2021) Estimation of linear and cyclic electron flows in photosynthesis based on ^{13}C-metabolic flux analysis. *Journal of Bioscience and Bioengineering.* 131 (3), 277-282. doi: 10.1016/j.jbiosc.2020.11.002.
Yang, J., Xiang, N., Liu, Y., Guo, C., Li, C., Li, H., Cai, S., Dixon, R. & Wang, Y.-P. (2023) Organelle-dependent polyprotein designs enable stoichiometric expression of nitrogen fixation components targeted to mitochondria. *Proceedings of the National Academy of Sciences.* 120 (34). doi: 10.1073/pnas.2305142120.
Yang, J., Xie, X., Xiang, N., Tian, Z.-X., Dixon, R. & Wang, Y.-P. (2018) Polyprotein strategy for stoichiometric assembly of nitrogen fixation components for synthetic biology. *Proceedings of the National Academy of Sciences.* 115 (36), E8509-E8517. doi: 10.1073/pnas.1804992115.
Yanni, Y. G. & Dazzo, F. B. (2010) Enhancement of rice production using endophytic strains of Rhizobium leguminosarum bv. trifolii in extensive field inoculation trials within the Egypt Nile delta. *Plant and Soil.* 336 (1-2), 129-142. doi: 10.1007/s11104-010-0454-7.
Yunus, I. S. (2019) *Production of fatty acid-derived fuels and chemicals in cyanobacteria.* PhD Imperial College London, London.
Zandy, S. L., Doherty, J. M., Wibisono, N. D. & Gonzales, R. A. (2017) High sensitivity HPLC method for analysis of in vivo extracellular GABA using optimized fluorescence parameters for o-phthalaldehyde (OPA)/sulfite derivatives. *Journal of Chromatography B.* 1055-1056, 1-7. doi: 10.1016/J.JCHROMB.2017.04.003.
Zayed, O., Hewedy, O. A., Abdelmoteleb, A., Ali, M., Youssef, M. S., Roumia, A. F., Seymour, D. & Yuan, Z.-C. (2023) Nitrogen Journey in Plants: From Uptake to Metabolism, Stress Response, and Microbe Interaction. *Biomolecules.* 13 (10).
Zehr, J. P., Jenkins, B. D., Short, S. M. & Steward, G. F. (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology.* 5 (7), 539-554. doi: 10.1046/j.1462-2920.2003.00451.x.
Zwietering, M. H., Jongenburger, I., Rombouts, F. M. & van't Riet, K. (1990) Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology.* 56 (6), 1875-1881.

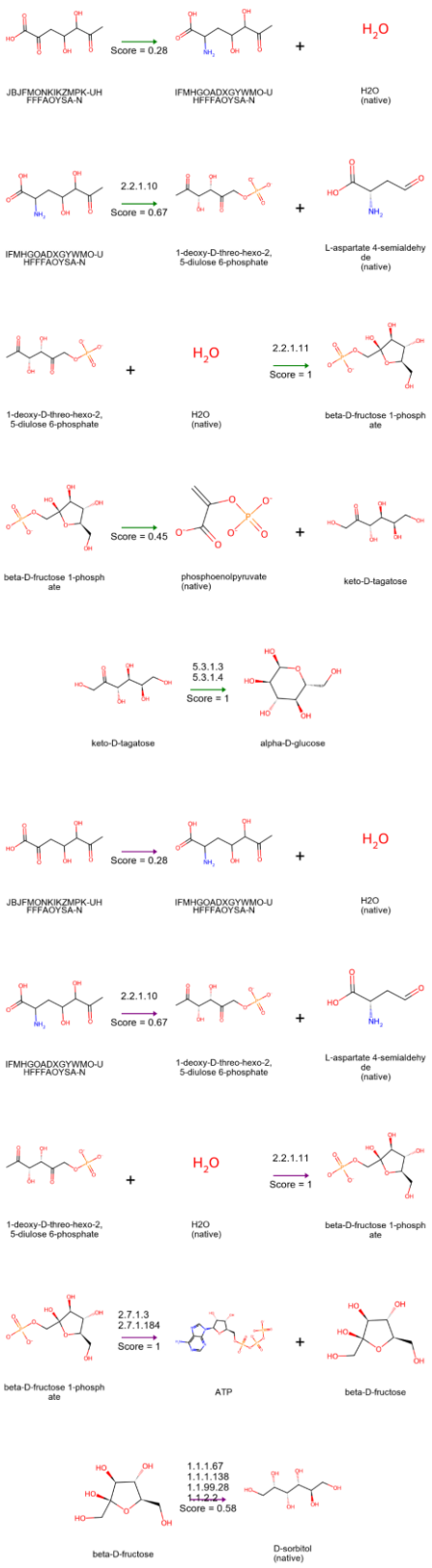
Chapter 9 Appendices

9.1 Supplementary Data

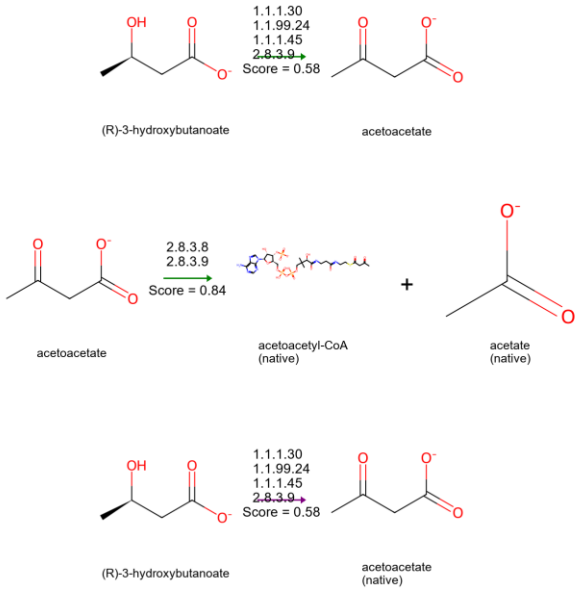
9.1.1 Exemplary Carbon Exchange Pathways



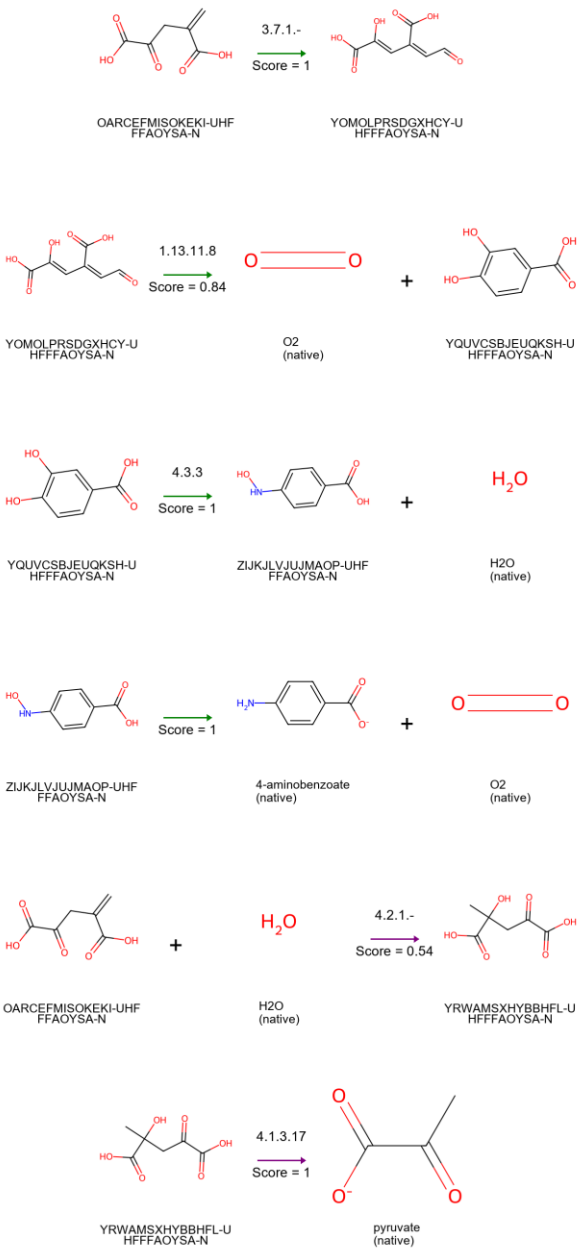
Pathway ID: C45584



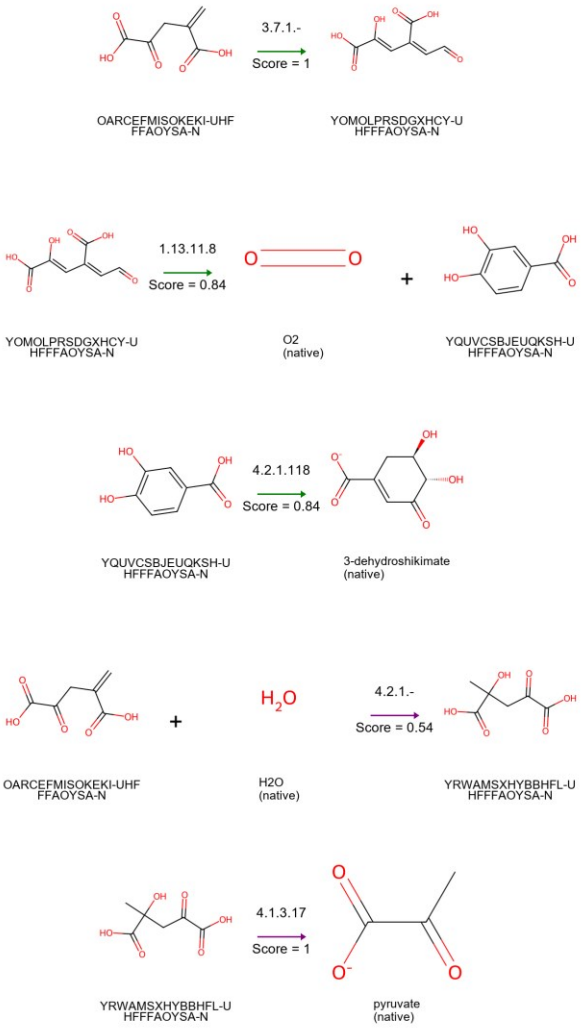
Pathway ID: C45662



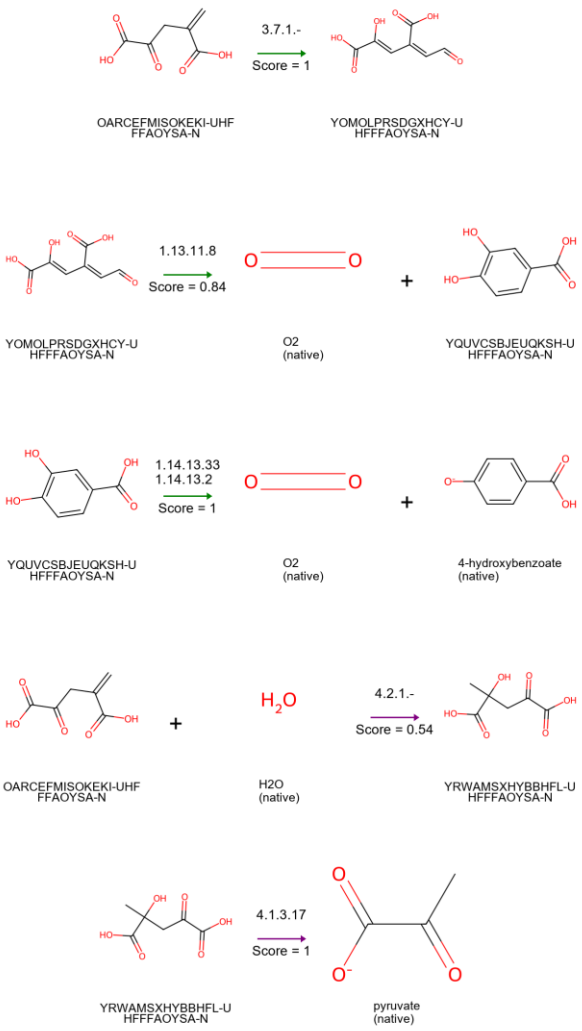
Pathway ID: C101263



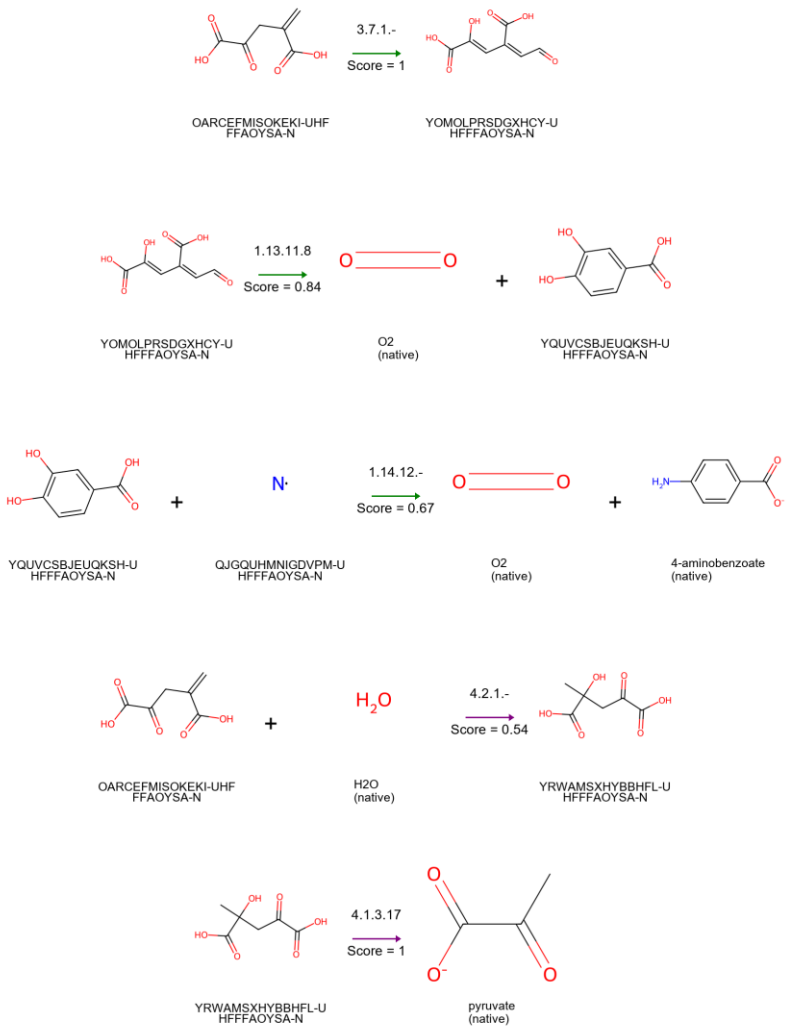
Pathway ID: C105505



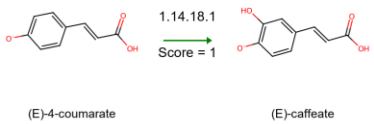
Pathway ID: C105522



Pathway ID: C105590



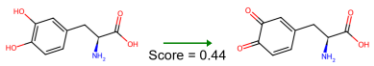
Pathway ID: C105692



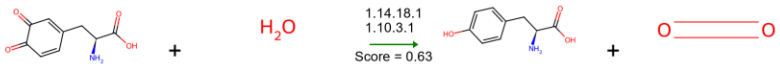
(E)-4-coumarate (E)-caffeate



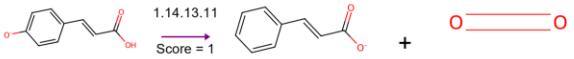
(E)-caffeate L-dopa



L-dopa L-dopaquinone

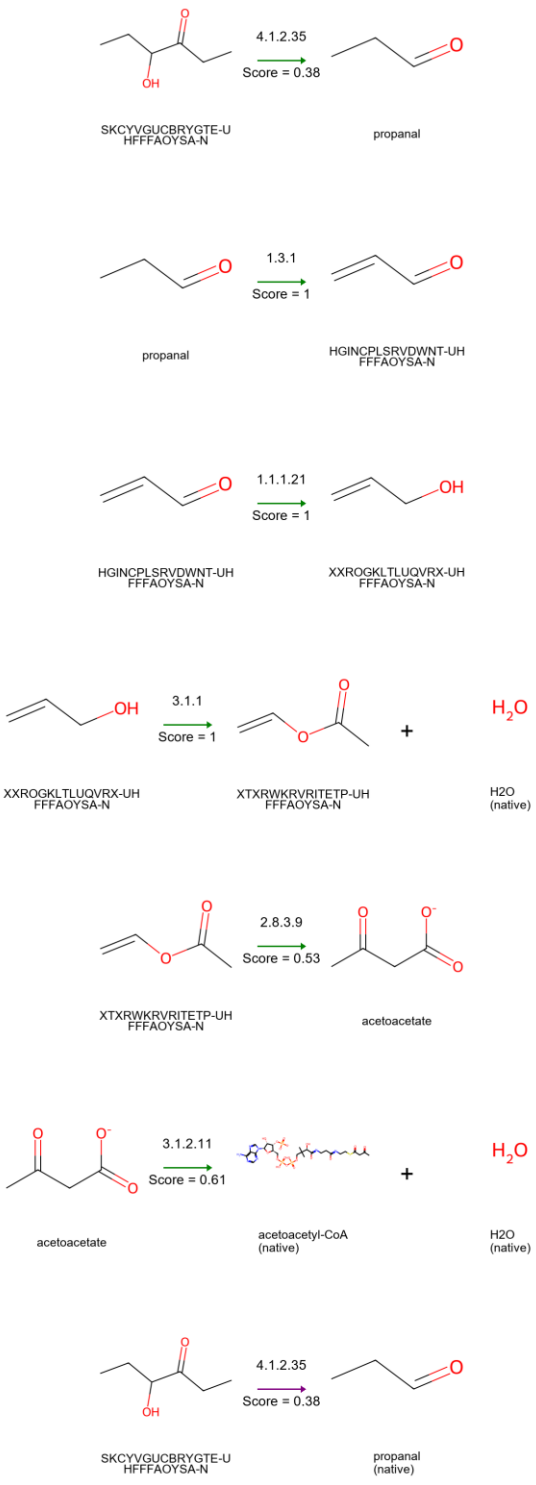


L-dopaquinone H2O (native) L-tyrosine (native) O2 (native)

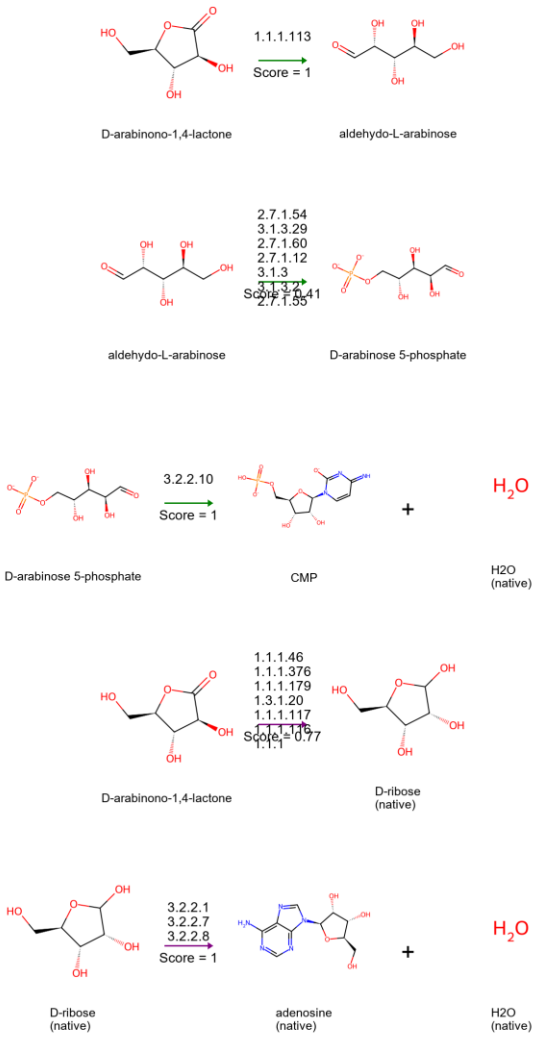


(E)-4-coumarate (E)-cinnamate O2 (native)

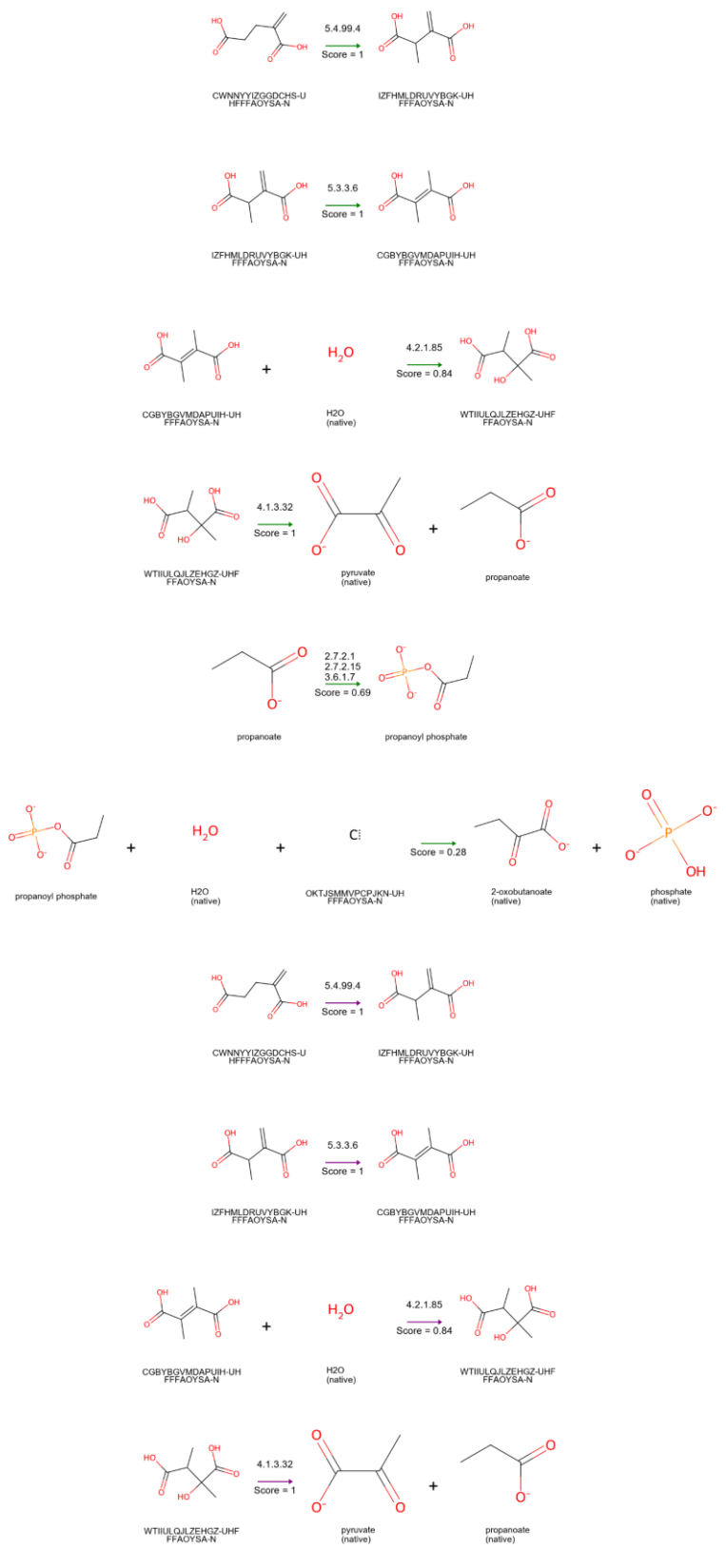
Pathway ID: C106024

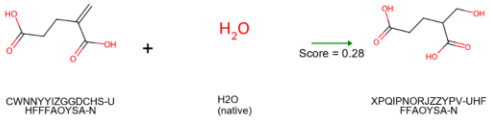


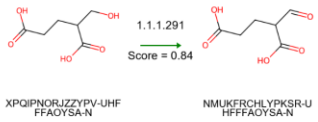
Pathway ID: C107013

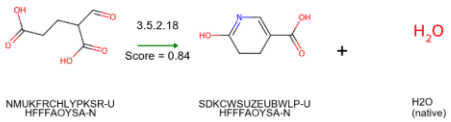


Pathway ID: C108795

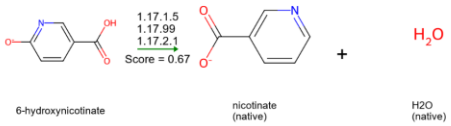

5.4.99.4
Score = 1
CWNNYYIZGGDCHS-UHFFFAOYSA-N
IZFHMLDRUVYBGK-UHFFFAOYSA-N
5.3.3.6
Score = 1
IZFHMLDRUVYBGK-UHFFFAOYSA-N
CGBYBGVMDAPUIH-UHFFFAOYSA-N
CGBYBGVMDAPUIH-UHFFFAOYSA-N
H2O
(native)
4.2.1.85
Score = 0.84
WTIIULQJLZEHGZ-UHFFFAOYSA-N
WTIIULQJLZEHGZ-UHFFFAOYSA-N
4.1.3.32
Score = 1
pyruvate
(native)
propanoate
propanoate
2.7.2.1
2.7.2.15
3.6.1.7
Score = 0.69
propanoyl phosphate
propanoyl phosphate
H2O
(native)
OKTJSMMVPCPJKN-UHFFFAOYSA-N
Score = 0.28
2-oxobutanoate
(native)
phosphate
(native)
5.4.99.4
Score = 1
CWNNYYIZGGDCHS-UHFFFAOYSA-N
IZFHMLDRUVYBGK-UHFFFAOYSA-N
5.3.3.6
Score = 1
IZFHMLDRUVYBGK-UHFFFAOYSA-N
CGBYBGVMDAPUIH-UHFFFAOYSA-N
CGBYBGVMDAPUIH-UHFFFAOYSA-N
H2O
(native)
4.2.1.85
Score = 0.84
WTIIULQJLZEHGZ-UHFFFAOYSA-N
WTIIULQJLZEHGZ-UHFFFAOYSA-N
4.1.3.32
Score = 1
pyruvate
(native)
propanoate
(native)


CWNNYYIZGGDCHS-U
HFFFAOYSA-N
H2O
H2O
(native)
Score = 0.28
XPQIPNORJZZYPV-UHF
FFAOYSA-N

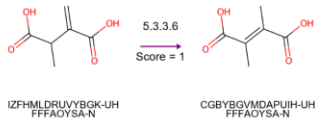

1.1.1.291
Score = 0.84
XPQIPNORJZZYPV-UHF
FFAOYSA-N
NMUKFRCHLYPKSR-U
HFFFAOYSA-N

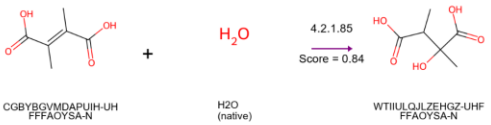

3.5.2.18
Score = 0.84
NMUKFRCHLYPKSR-U
HFFFAOYSA-N
SDKCWSUZEUBWLP-U
HFFFAOYSA-N
H2O
(native)

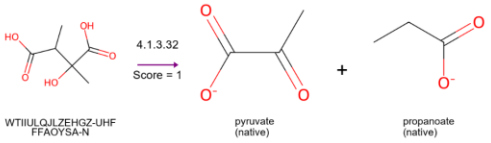

1.3.7.1
Score = 0.84
SDKCWSUZEUBWLP-U
HFFFAOYSA-N
6-hydroxynicotinate

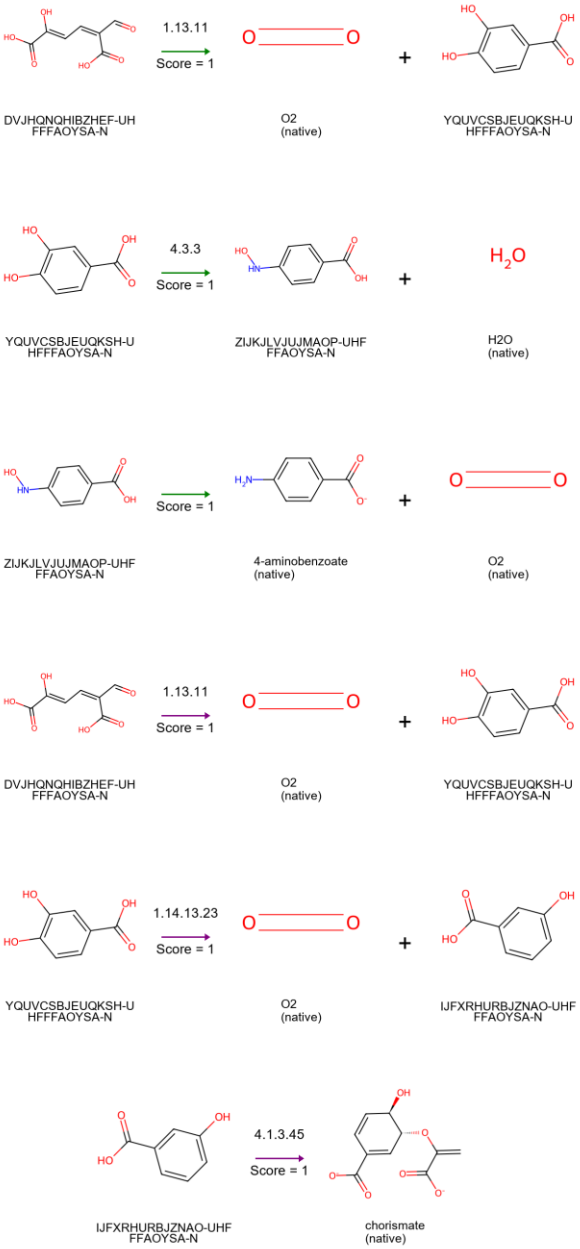

1.17.1.5
1.17.99
1.17.2.1
Score = 0.67
6-hydroxynicotinate
nicotinate
(native)
H2O
(native)


5.4.99.4
Score = 1
CWNNYYIZGGDCHS-U
HFFFAOYSA-N
IZFHMLDRUVYBGK-UH
FFFAOYSA-N

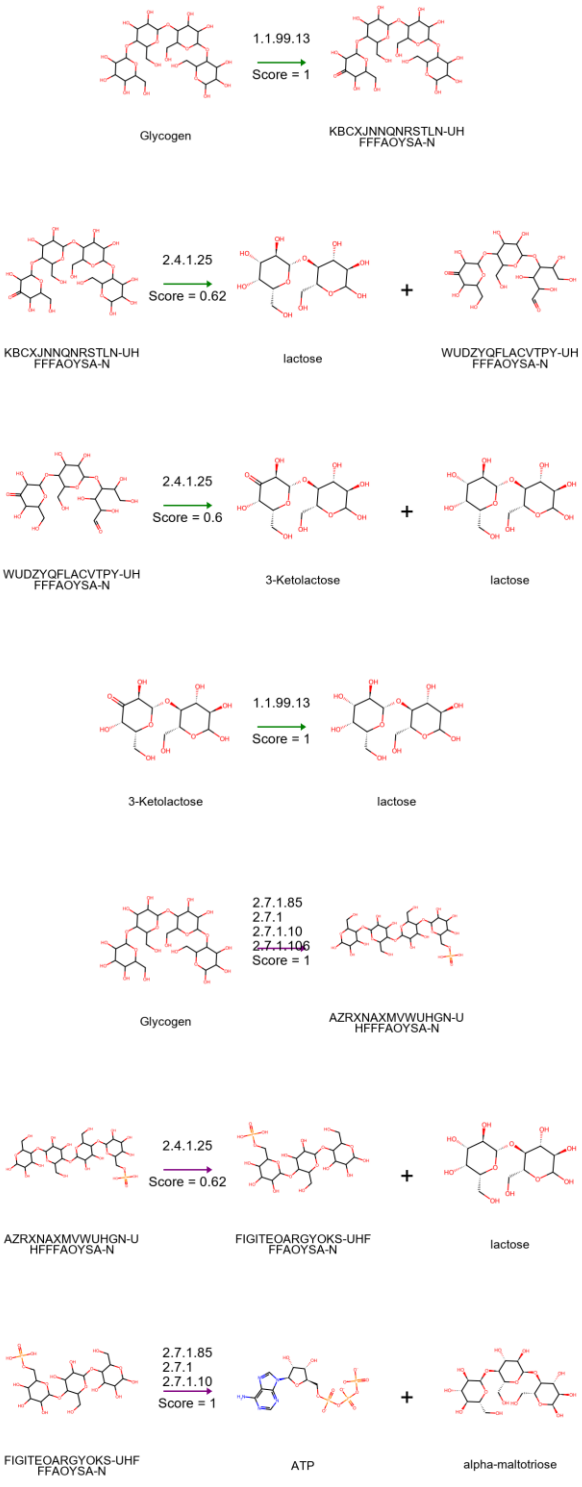

5.3.3.6
Score = 1
IZFHMLDRUVYBGK-UH
FFFAOYSA-N
CGBYBGVMDAPUIH-UH
FFFAOYSA-N


CGBYBGVMDAPUIH-UH
FFFAOYSA-N
H2O
(native)
4.2.1.85
Score = 0.84
WTIIULQJLZEHGZ-UHF
FFAOYSA-N

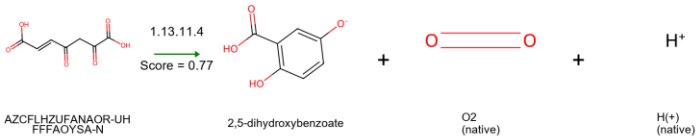

4.1.3.32
Score = 1
WTIIULQJLZEHGZ-UHF
FFAOYSA-N
pyruvate
(native)
propanoate
(native)

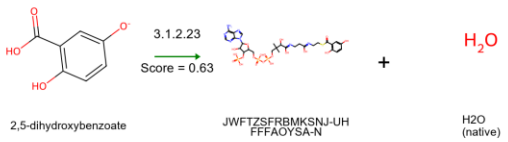


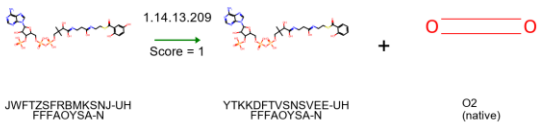
Pathway ID: C122489

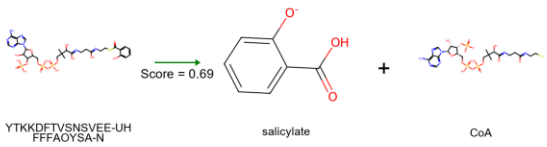


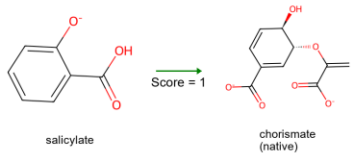
Pathway ID: C132249

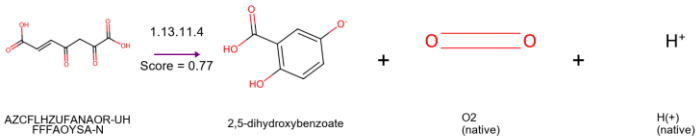

1.13.11.4
Score = 0.77
AZCFLHZUFANAOR-UH FFFAOYSA-N
2,5-dihydroxybenzoate
O2 (native)
H(+) (native)

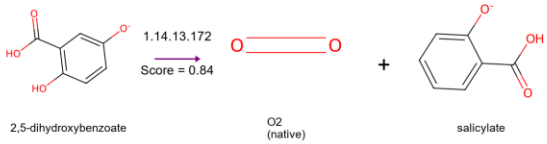

3.1.2.23
Score = 0.63
2,5-dihydroxybenzoate
JWFTZSFRBMKSNJ-UH FFFAOYSA-N
H2O (native)

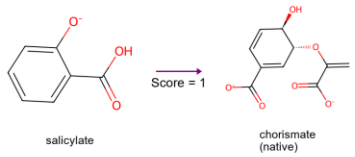

1.14.13.209
Score = 1
JWFTZSFRBMKSNJ-UH FFFAOYSA-N
YTKKDFTVSNSVEE-UH FFFAOYSA-N
O2 (native)

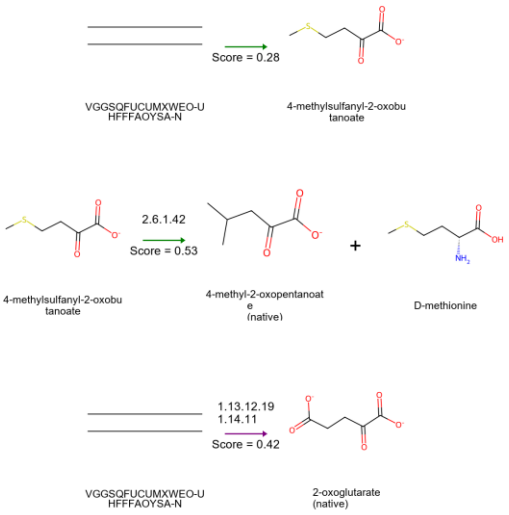

Score = 0.69
YTKKDFTVSNSVEE-UH FFFAOYSA-N
salicylate
CoA


Score = 1
salicylate
chorismate (native)

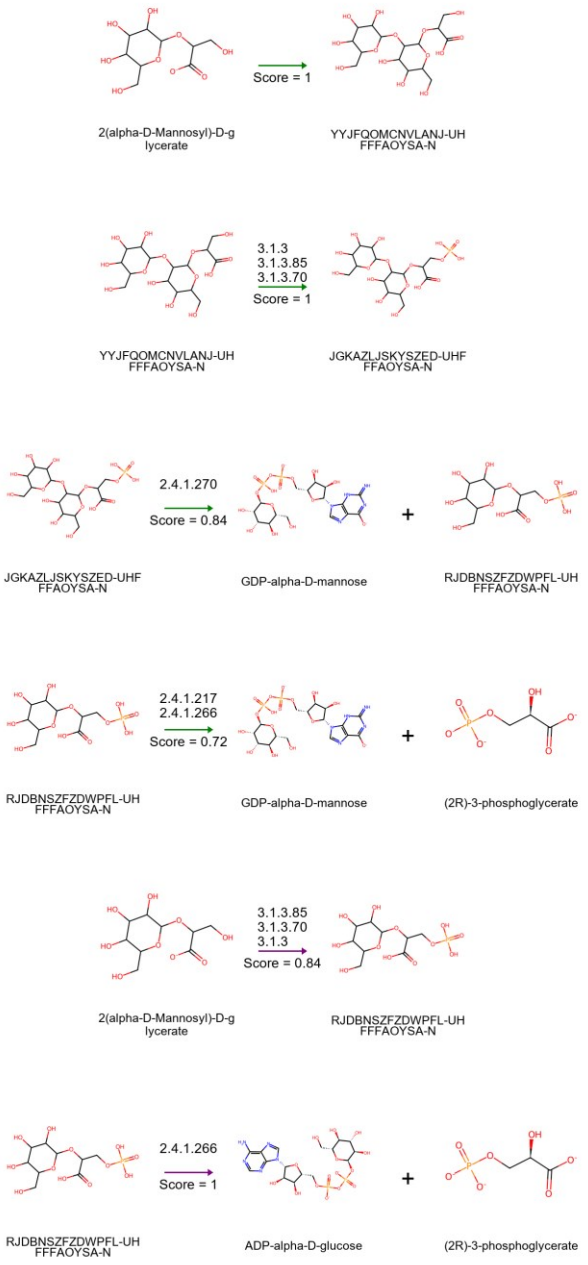

1.13.11.4
Score = 0.77
AZCFLHZUFANAOR-UH FFFAOYSA-N
2,5-dihydroxybenzoate
O2 (native)
H(+) (native)


1.14.13.172
Score = 0.84
2,5-dihydroxybenzoate
O2 (native)
salicylate

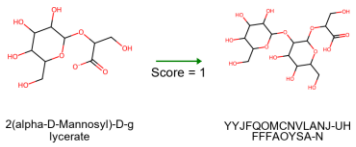

Score = 1
salicylate
chorismate (native)

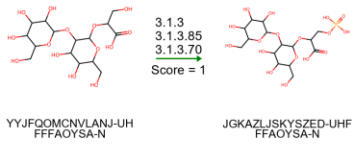


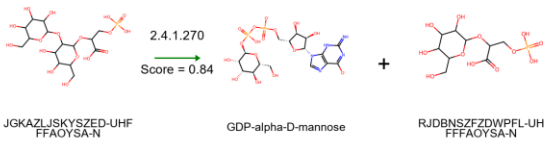
Pathway ID: C137485

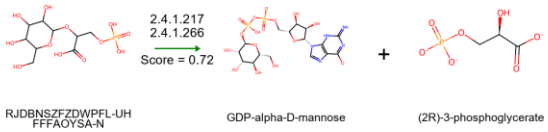


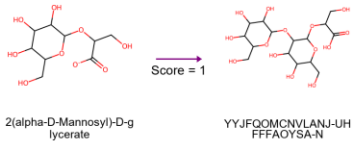
Pathway ID: C146421

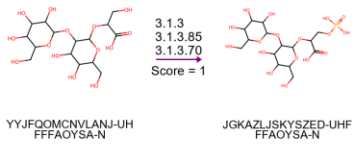

Score = 1
2(alpha-D-Mannosyl)-D-glycerate
YYJFQOMCNVLANJ-UHFFFAOYSA-N

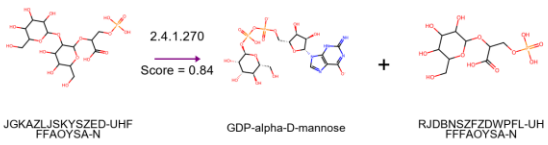

3.1.3
3.1.3.85
3.1.3.70
Score = 1
YYJFQOMCNVLANJ-UHFFFAOYSA-N
JGKAZLJSKYSZED-UHFFFAOYSA-N

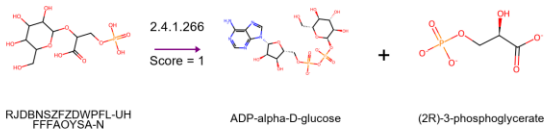

2.4.1.270
Score = 0.84
JGKAZLJSKYSZED-UHFFFAOYSA-N
GDP-alpha-D-mannose
RJDBNSZFZDWPFL-UHFFFAOYSA-N

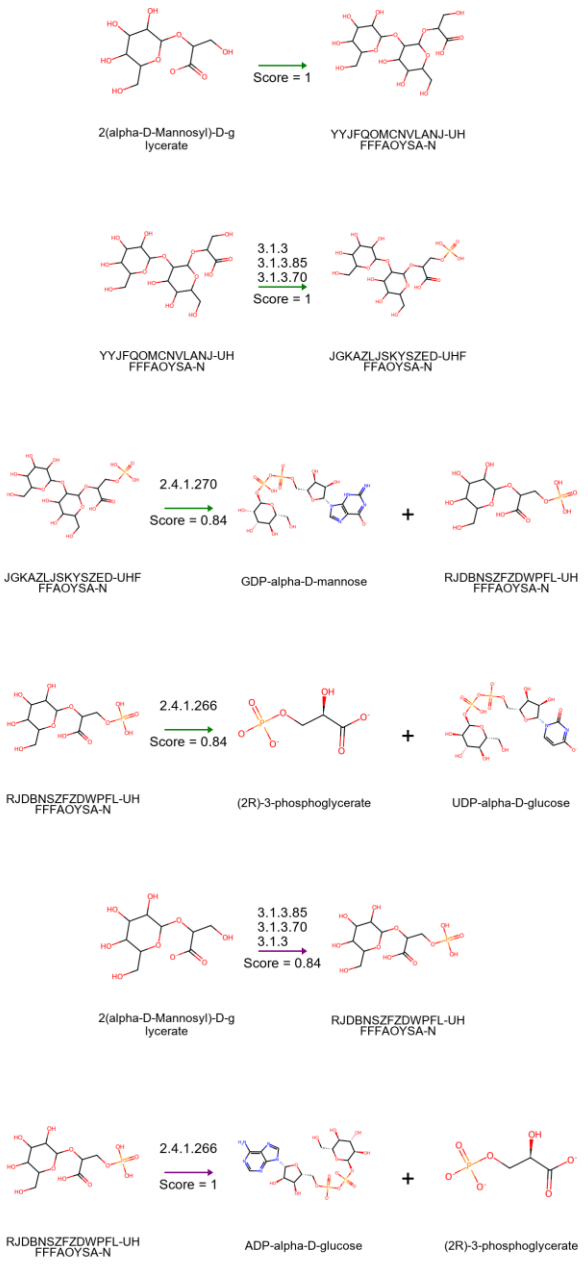

2.4.1.217
2.4.1.266
Score = 0.72
RJDBNSZFZDWPFL-UHFFFAOYSA-N
GDP-alpha-D-mannose
(2R)-3-phosphoglycerate


Score = 1
2(alpha-D-Mannosyl)-D-glycerate
YYJFQOMCNVLANJ-UHFFFAOYSA-N

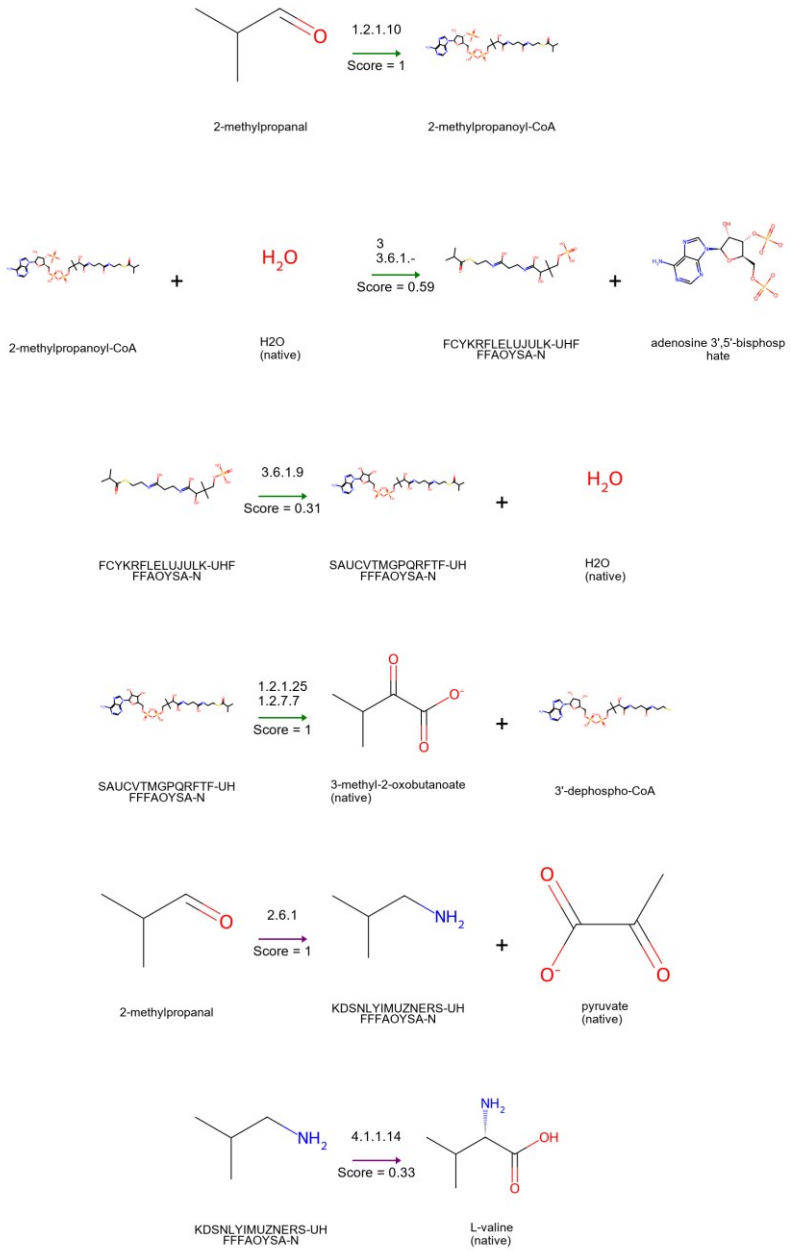

3.1.3
3.1.3.85
3.1.3.70
Score = 1
YYJFQOMCNVLANJ-UHFFFAOYSA-N
JGKAZLJSKYSZED-UHFFFAOYSA-N


2.4.1.270
Score = 0.84
JGKAZLJSKYSZED-UHFFFAOYSA-N
GDP-alpha-D-mannose
RJDBNSZFZDWPFL-UHFFFAOYSA-N

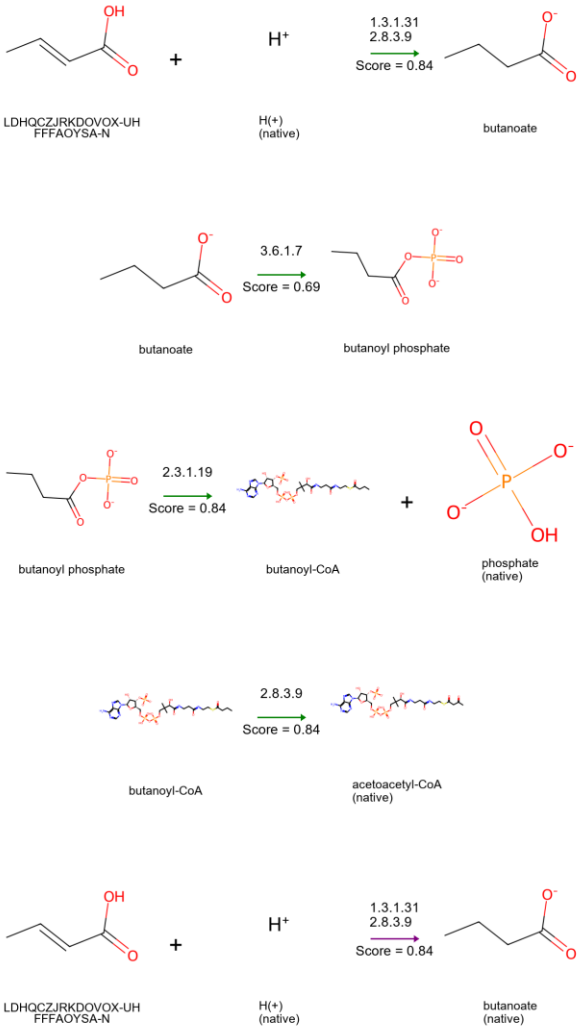

2.4.1.266
Score = 1
RJDBNSZFZDWPFL-UHFFFAOYSA-N
ADP-alpha-D-glucose
(2R)-3-phosphoglycerate



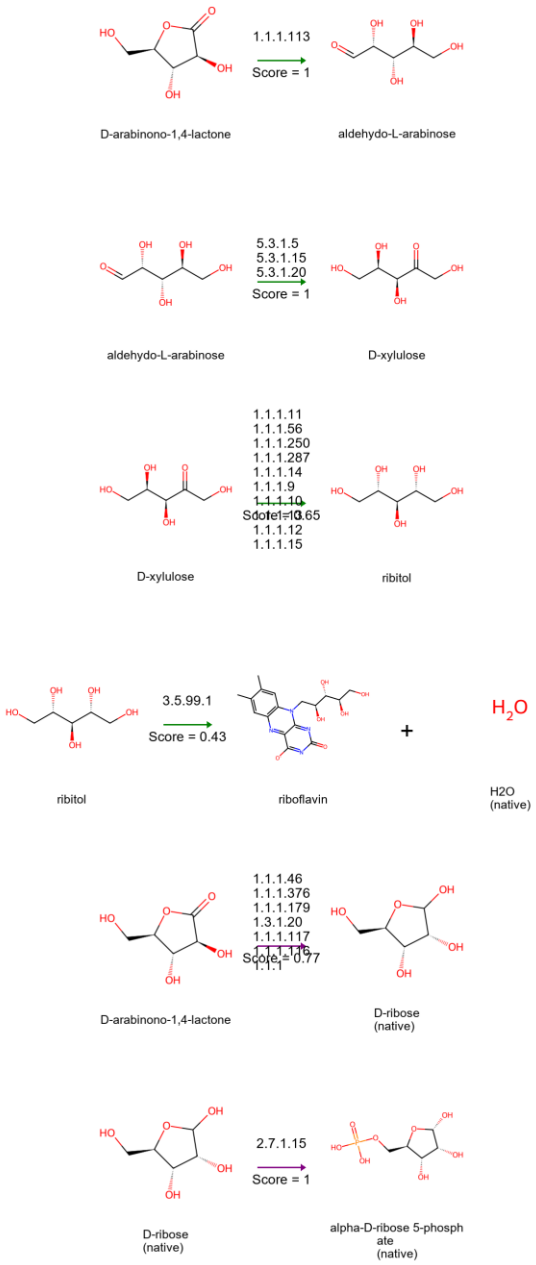
Pathway ID: C146461



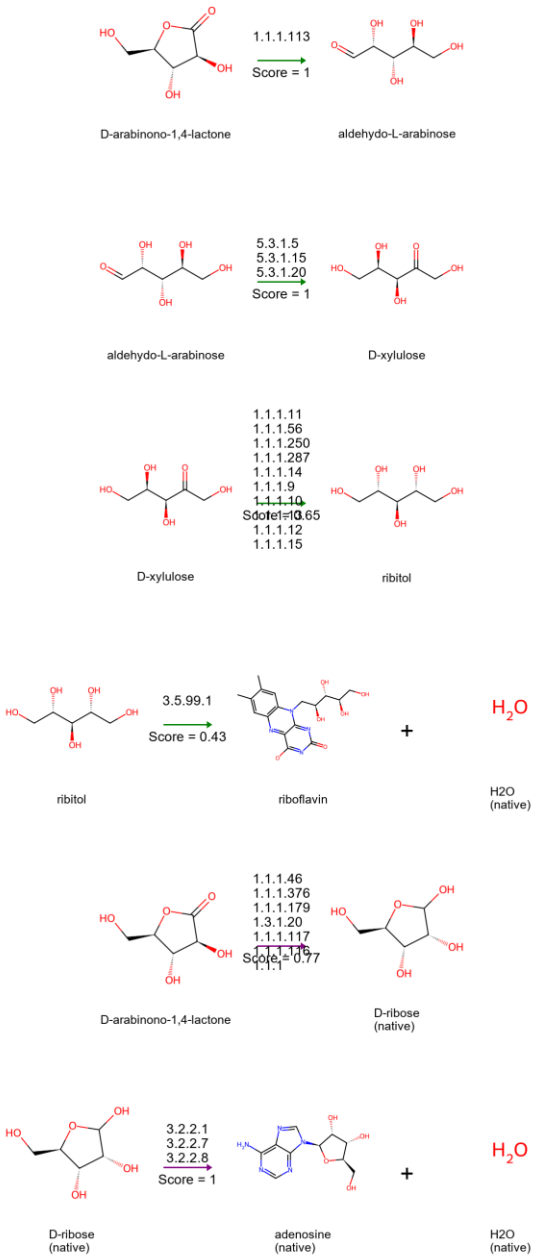
Pathway ID: C155050



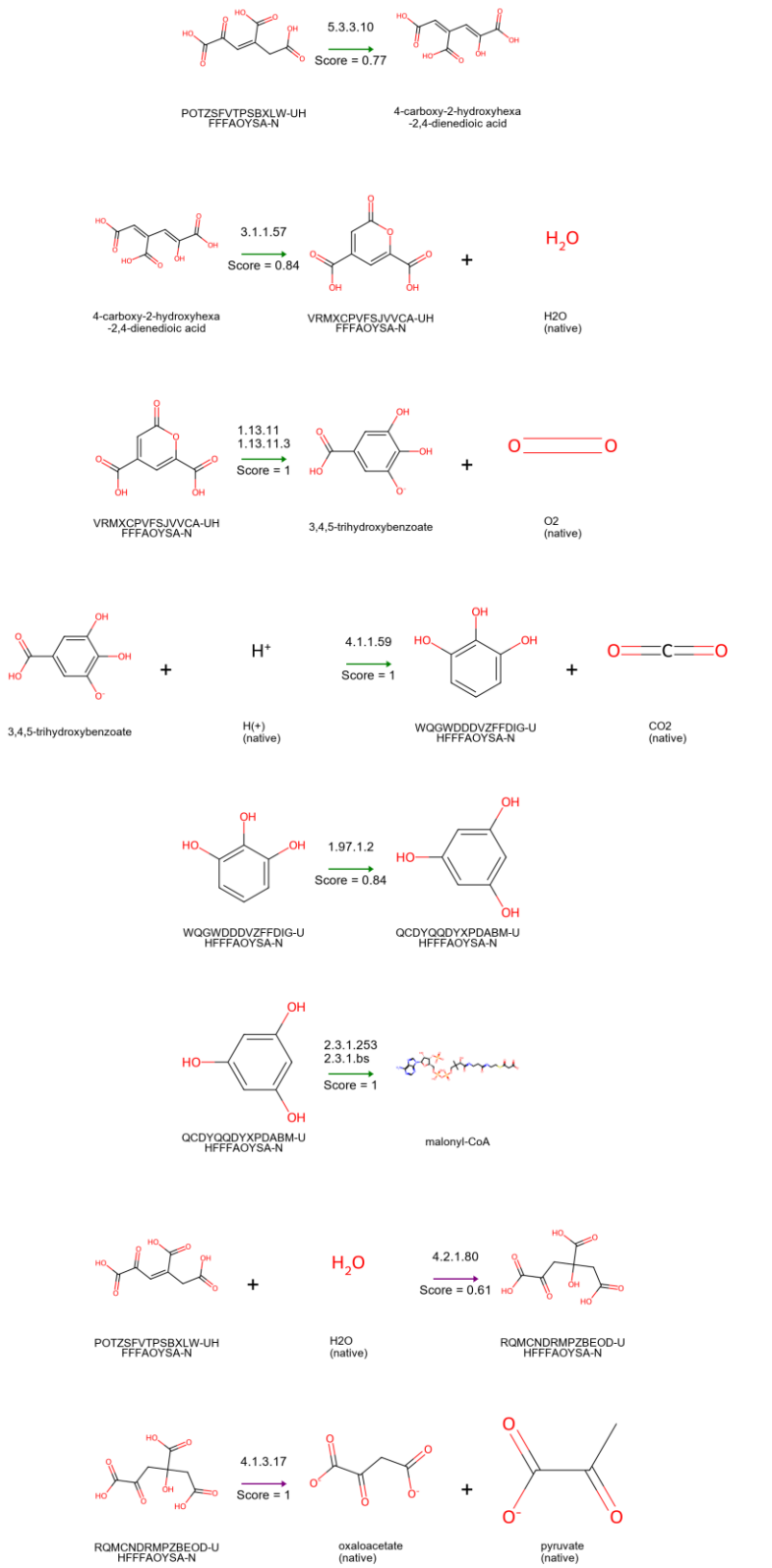
Pathway ID: C155695



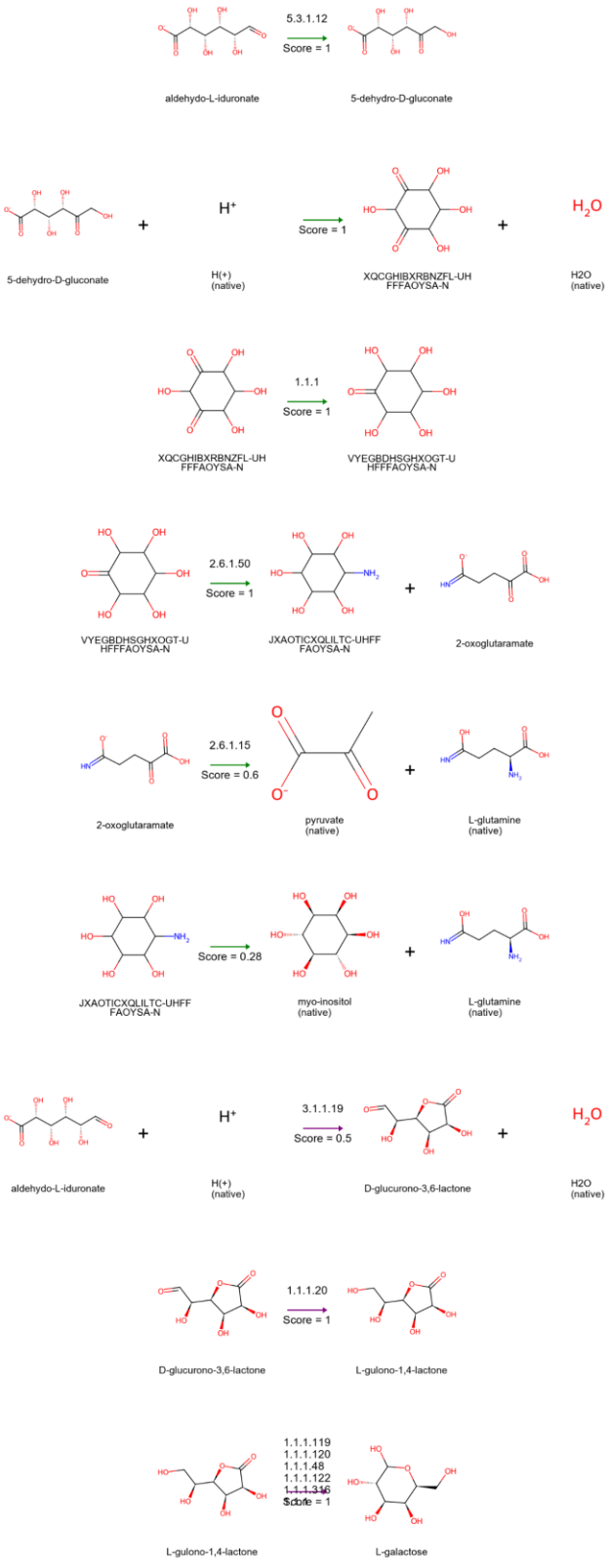
Pathway ID: C173913



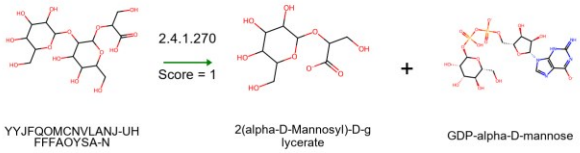
Pathway ID: C173915

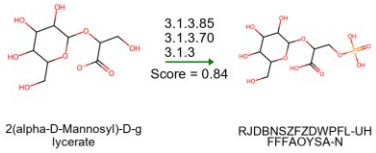


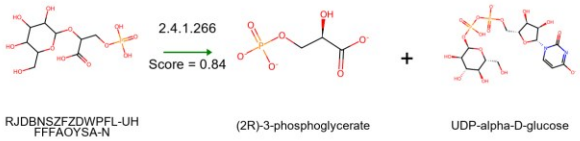
Pathway ID: C177917

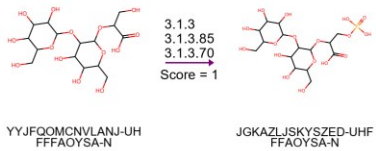


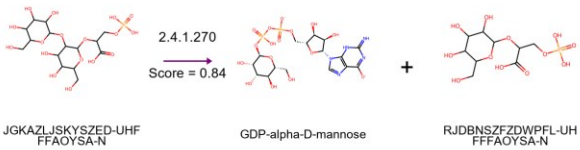
Pathway ID: C180454

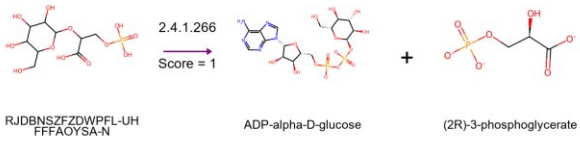




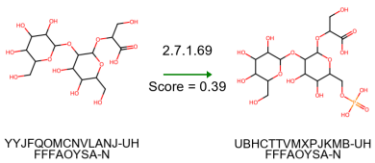


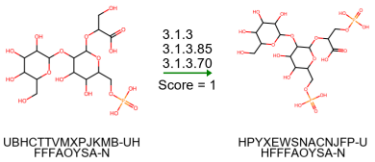


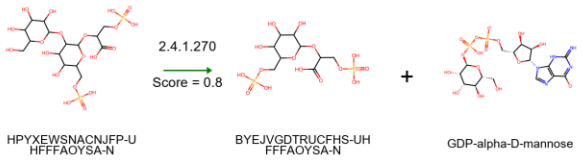


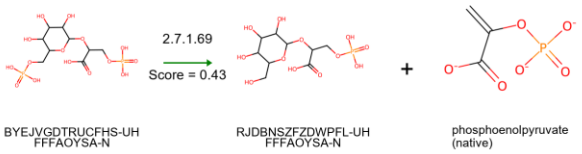


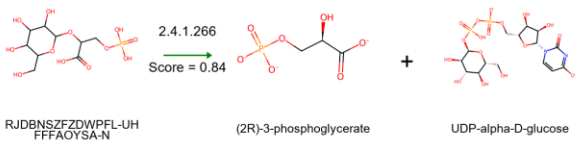
Pathway ID: C181313

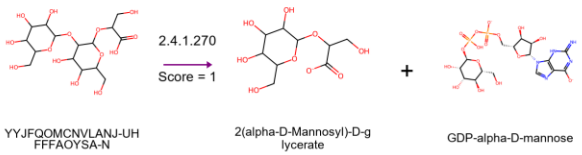


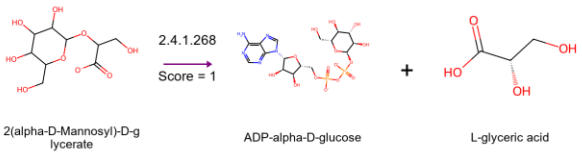




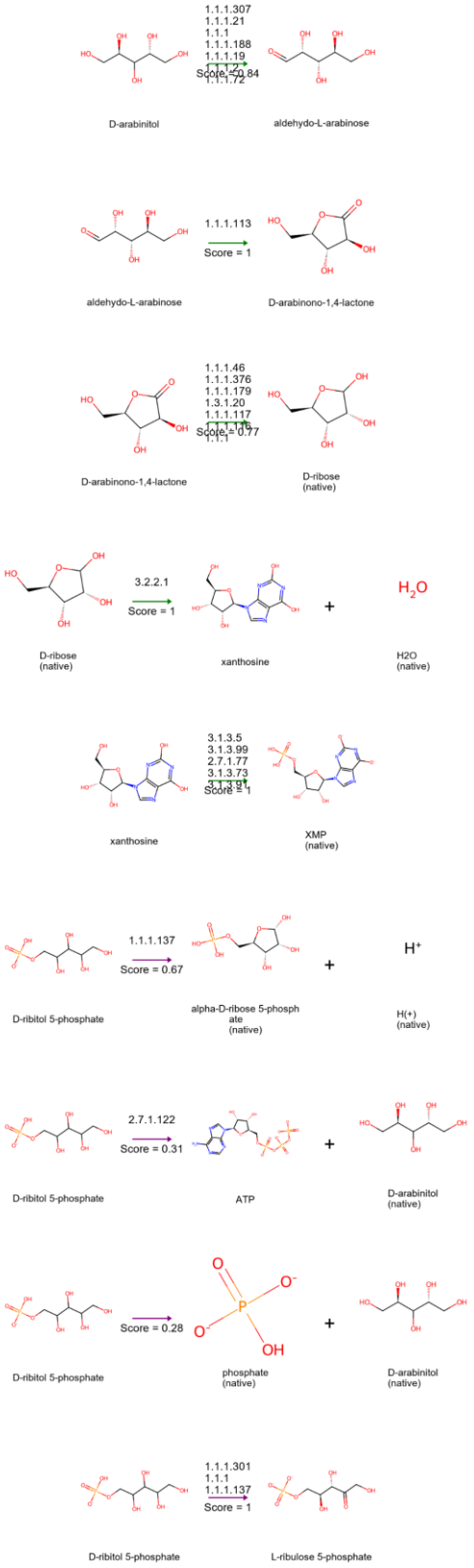


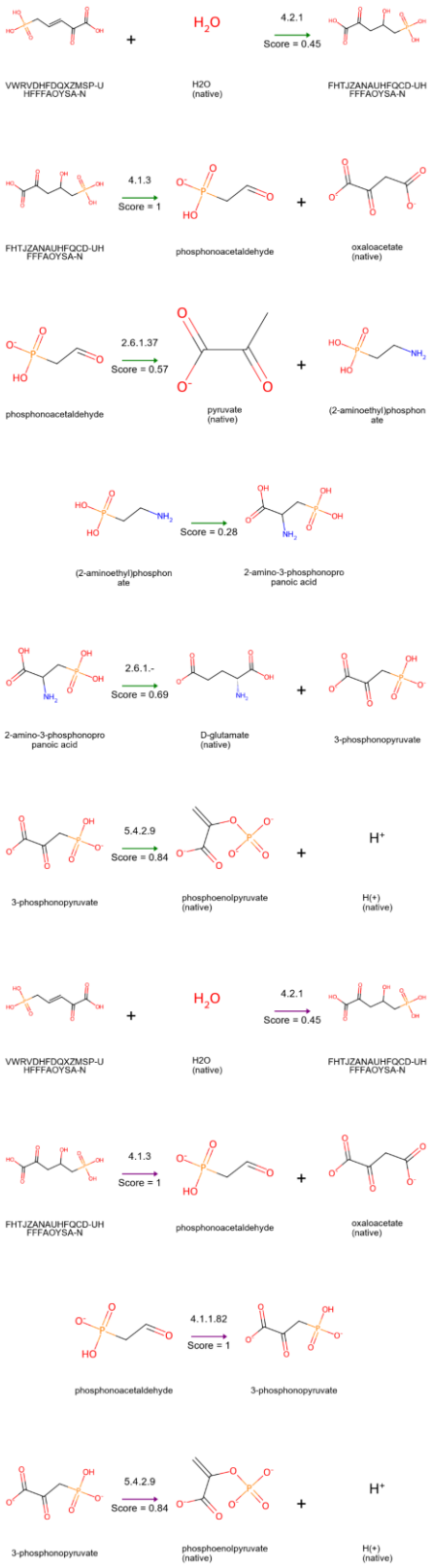


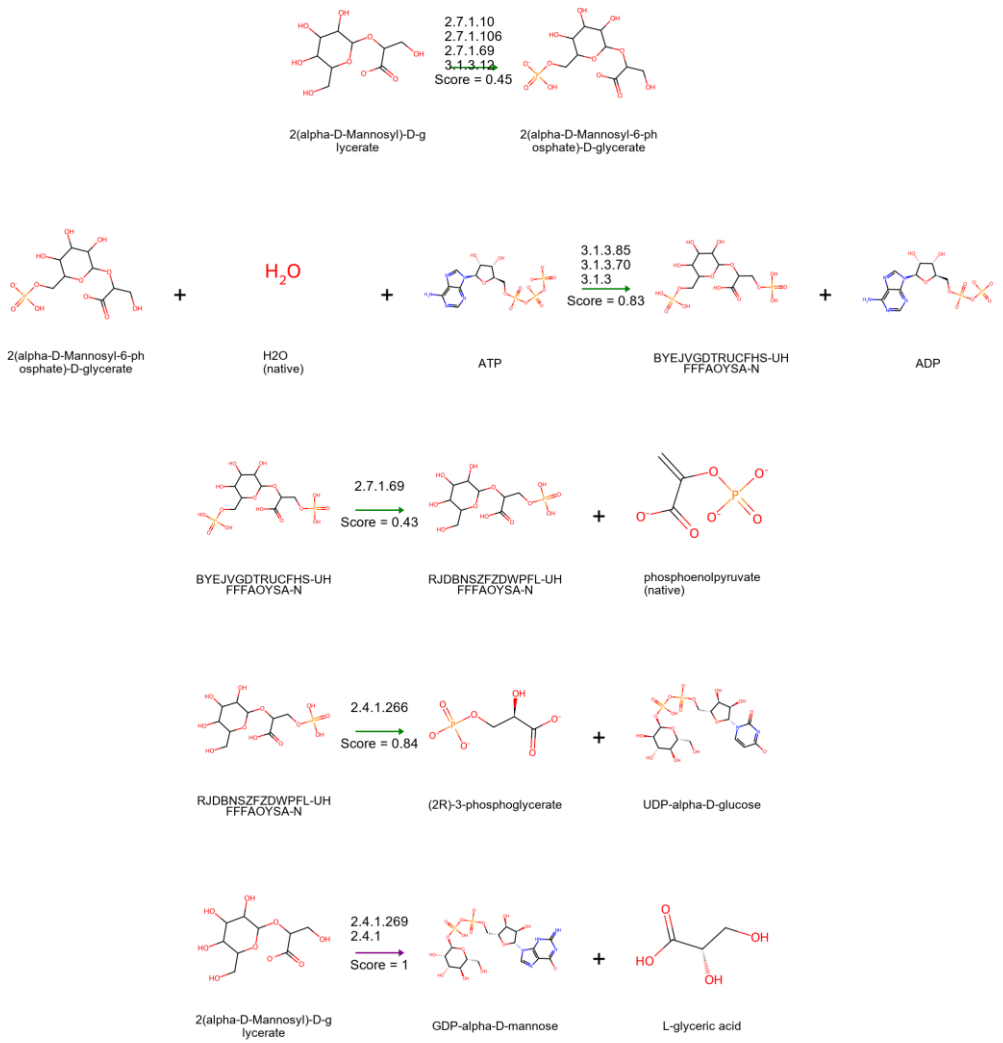




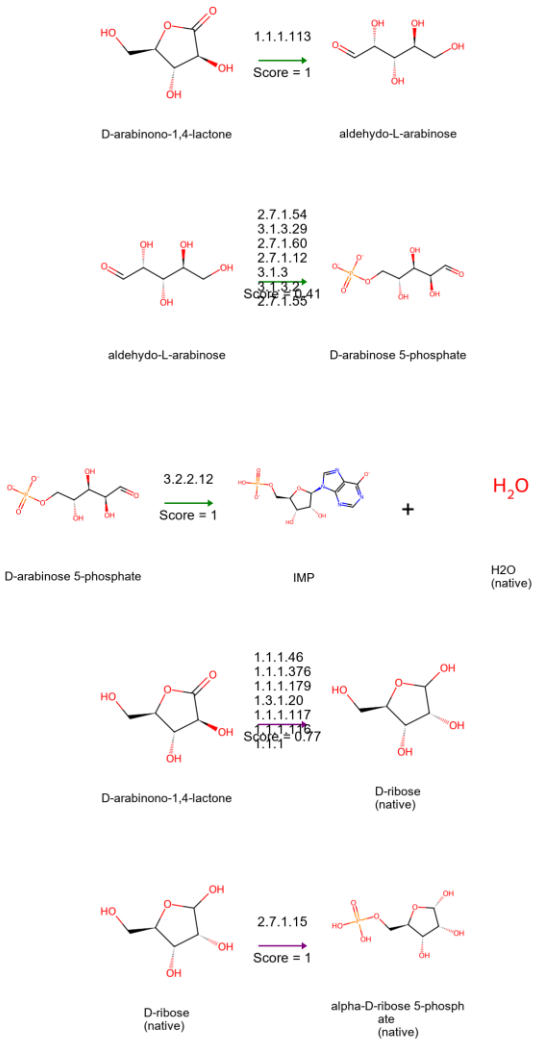
Pathway ID: C181479


1.1.1.307
1.1.1.21
1.1.1
1.1.1.188
1.1.1.19
1.1.1.2
Score = 0.84
1.1.1.72
D-arabinitol
aldehydo-L-arabinose
1.1.1.113
Score = 1
aldehydo-L-arabinose
D-arabinono-1,4-lactone
1.1.1.46
1.1.1.376
1.1.1.179
1.3.1.20
1.1.1.117
1.1.1.119
Score = 0.77
1.1.1
D-arabinono-1,4-lactone
D-ribose
(native)
3.2.2.1
Score = 1
D-ribose
(native)
xanthosine
H2O
(native)
3.1.3.5
3.1.3.99
2.7.1.77
3.1.3.73
3.1.3.91
Score = 1
xanthosine
XMP
(native)
1.1.1.137
Score = 0.67
D-ribitol 5-phosphate
alpha-D-ribose 5-phosph
ate
(native)
H(+)
(native)
2.7.1.122
Score = 0.31
D-ribitol 5-phosphate
ATP
D-arabinitol
(native)
Score = 0.28
D-ribitol 5-phosphate
phosphate
(native)
D-arabinitol
(native)
1.1.1.301
1.1.1
1.1.1.137
Score = 1
D-ribitol 5-phosphate
L-ribulose 5-phosphate

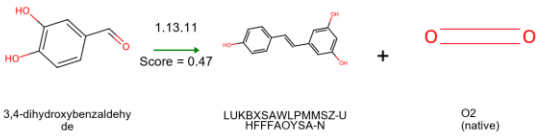

VWRVDHFDQXZMSP-UHFFFAOYSA-N
H2O (native)
FHTJZANAUHFQCD-UHFFFAOYSA-N
4.2.1
Score = 0.45
4.1.3
Score = 1
phosphonoacetaldehyde
oxaloacetate (native)
2.6.1.37
Score = 0.57
pyruvate (native)
(2-aminoethyl)phosphonate
Score = 0.28
2-amino-3-phosphonopropanoic acid
2.6.1.-
Score = 0.69
D-glutamate (native)
3-phosphonopyruvate
5.4.2.9
Score = 0.84
phosphoenolpyruvate (native)
H(+) (native)
4.1.1.82
Score = 1

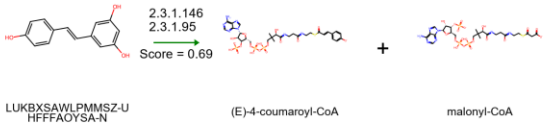


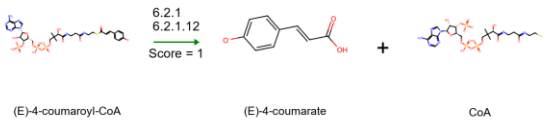
Pathway ID: C207651

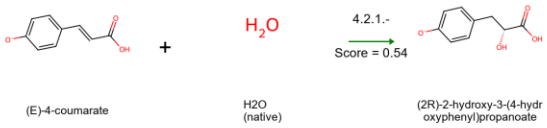


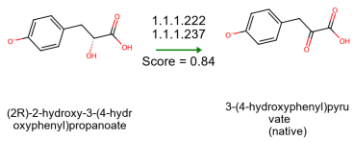
Pathway ID: C209823

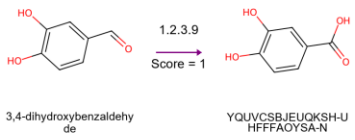

1.13.11
Score = 0.47
3,4-dihydroxybenzaldehyde
LUKBXSAWLPMMSZ-UHFFFAOYSA-N
O2 (native)


2.3.1.146
2.3.1.95
Score = 0.69
LUKBXSAWLPMMSZ-UHFFFAOYSA-N
(E)-4-coumaroyl-CoA
malonyl-CoA

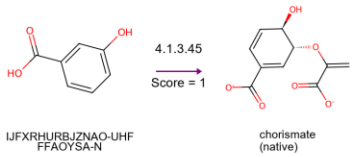

6.2.1
6.2.1.12
Score = 1
(E)-4-coumaroyl-CoA
(E)-4-coumarate
CoA

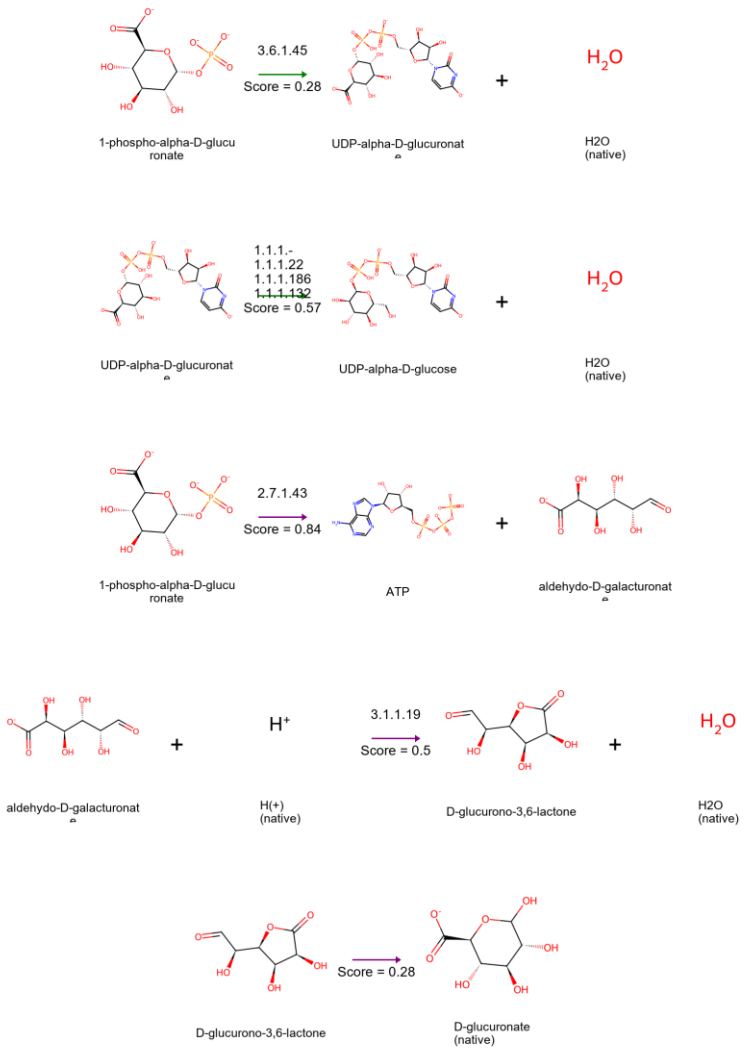

4.2.1.-
Score = 0.54
(E)-4-coumarate
H2O (native)
(2R)-2-hydroxy-3-(4-hydroxyphenyl)propanoate


1.1.1.222
1.1.1.237
Score = 0.84
(2R)-2-hydroxy-3-(4-hydroxyphenyl)propanoate
3-(4-hydroxyphenyl)pyruvate (native)

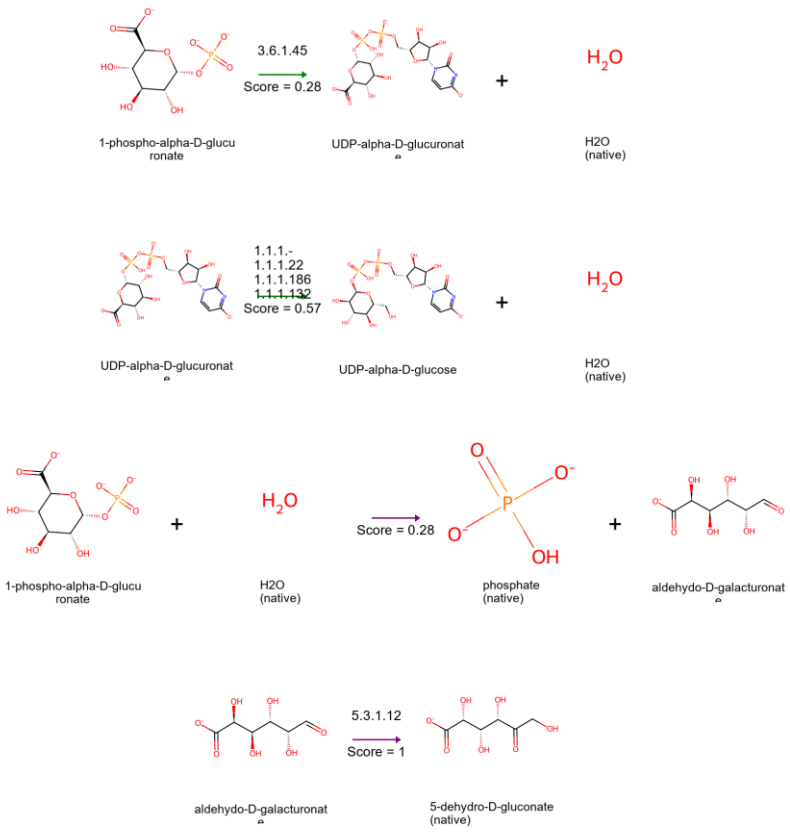

1.2.3.9
Score = 1
3,4-dihydroxybenzaldehyde
YQUVCSBJEUQKSH-UHFFFAOYSA-N


1.14.13.23
Score = 1
YQUVCSBJEUQKSH-UHFFFAOYSA-N
O2 (native)
IJFXRHURBJZNAO-UHFFFAOYSA-N

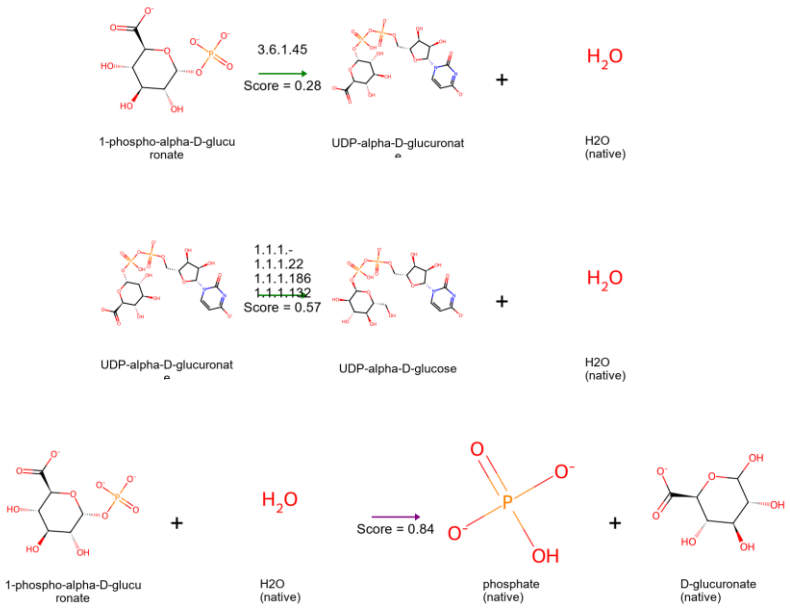

4.1.3.45
Score = 1
IJFXRHURBJZNAO-UHFFFAOYSA-N
chorismate (native)



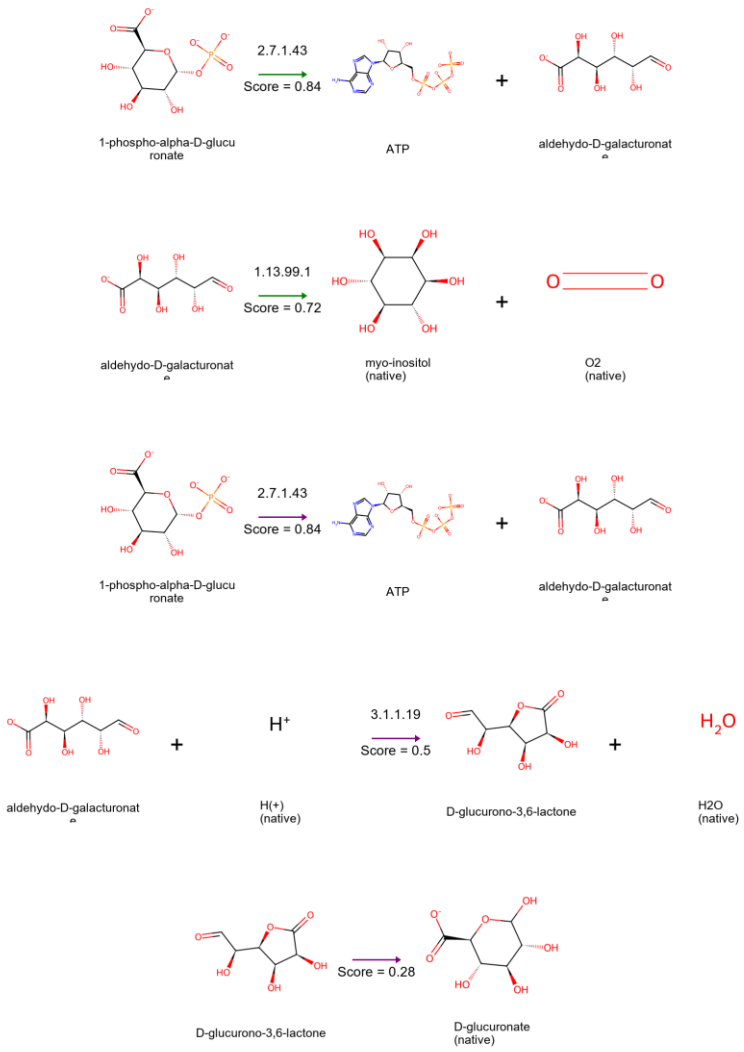
Pathway ID: C216637



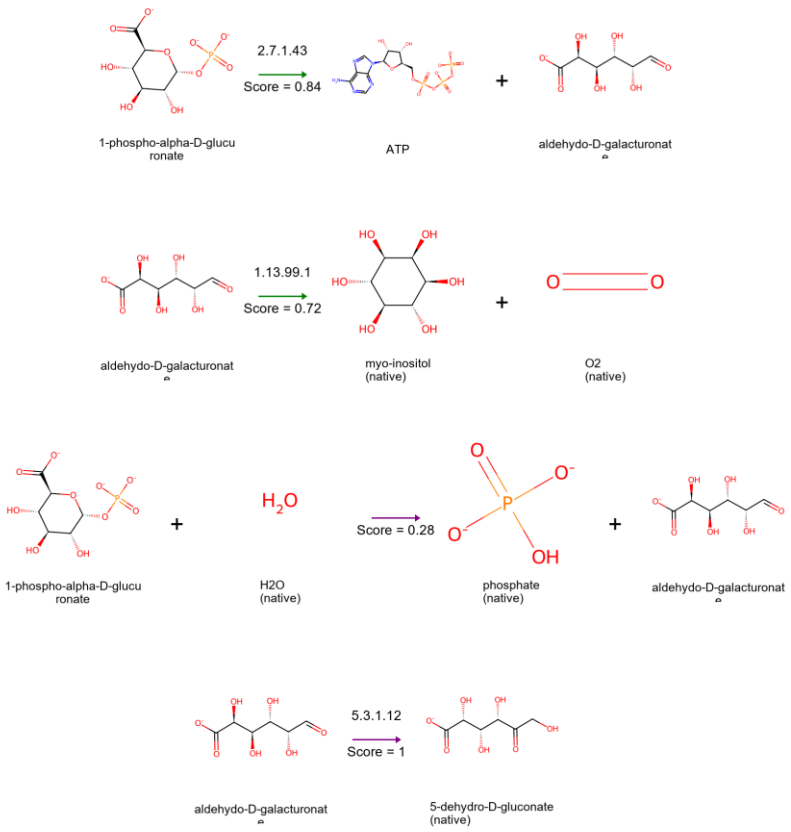
Pathway ID: C216639



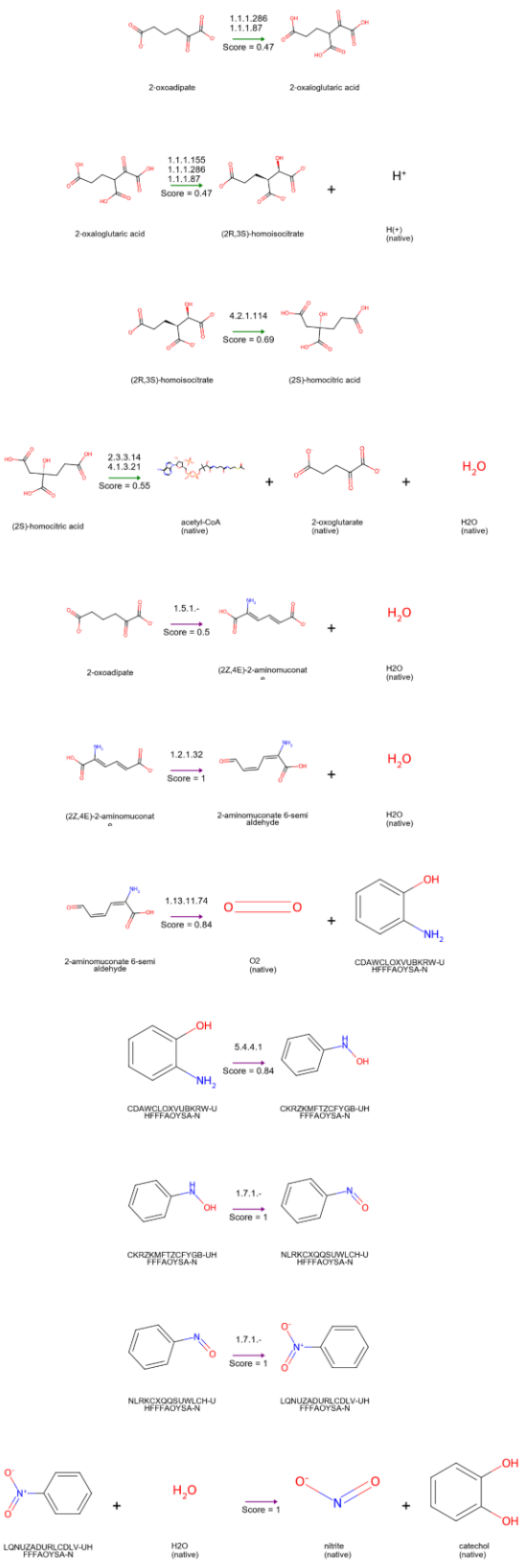
Pathway ID: C216642



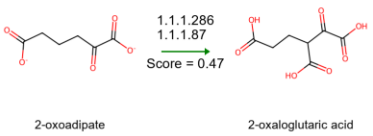
Pathway ID: C216651



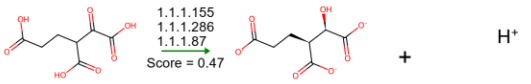
Pathway ID: C216653



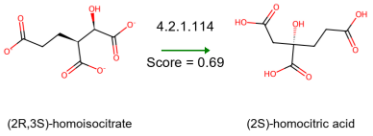
Pathway ID: C237451



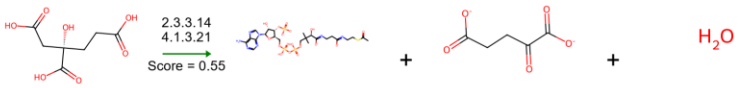
2-oxoadipate 2-oxaloglutaric acid



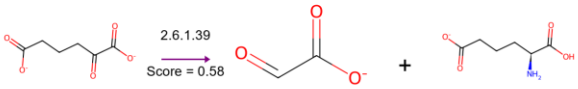
2-oxaloglutaric acid (2R,3S)-homoisocitrate H(+) (native)



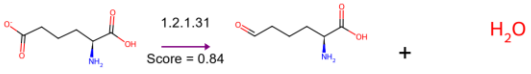
(2R,3S)-homoisocitrate (2S)-homocitric acid



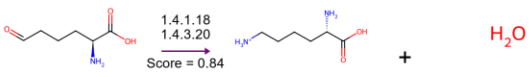
(2S)-homocitric acid acetyl-CoA (native) 2-oxoglutarate (native) H2O (native)



2-oxoadipate glyoxylate (native) L-2-aminoadipate

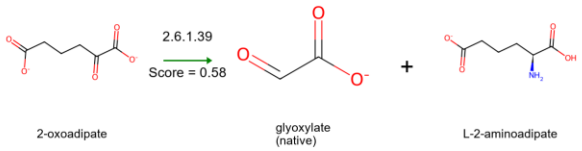


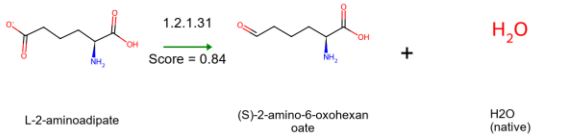
L-2-aminoadipate (S)-2-amino-6-oxohexanoate H2O (native)

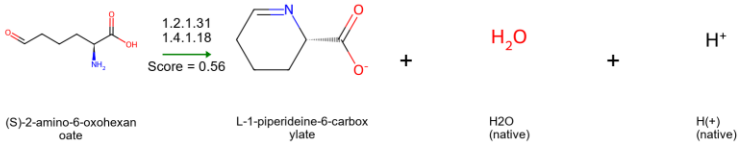


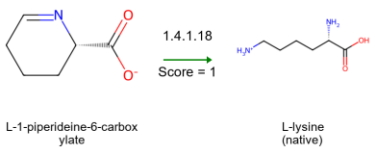
(S)-2-amino-6-oxohexanoate L-lysine (native) H2O (native)

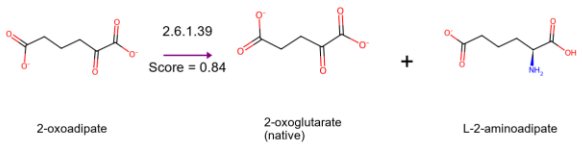
Pathway ID: C237455

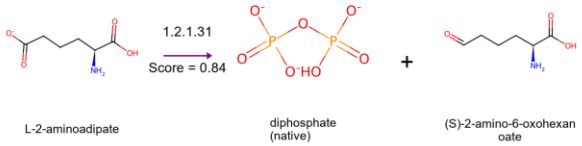


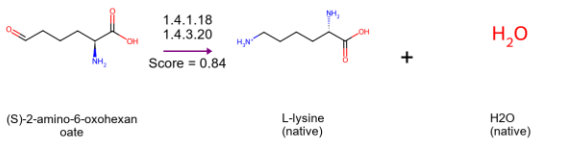




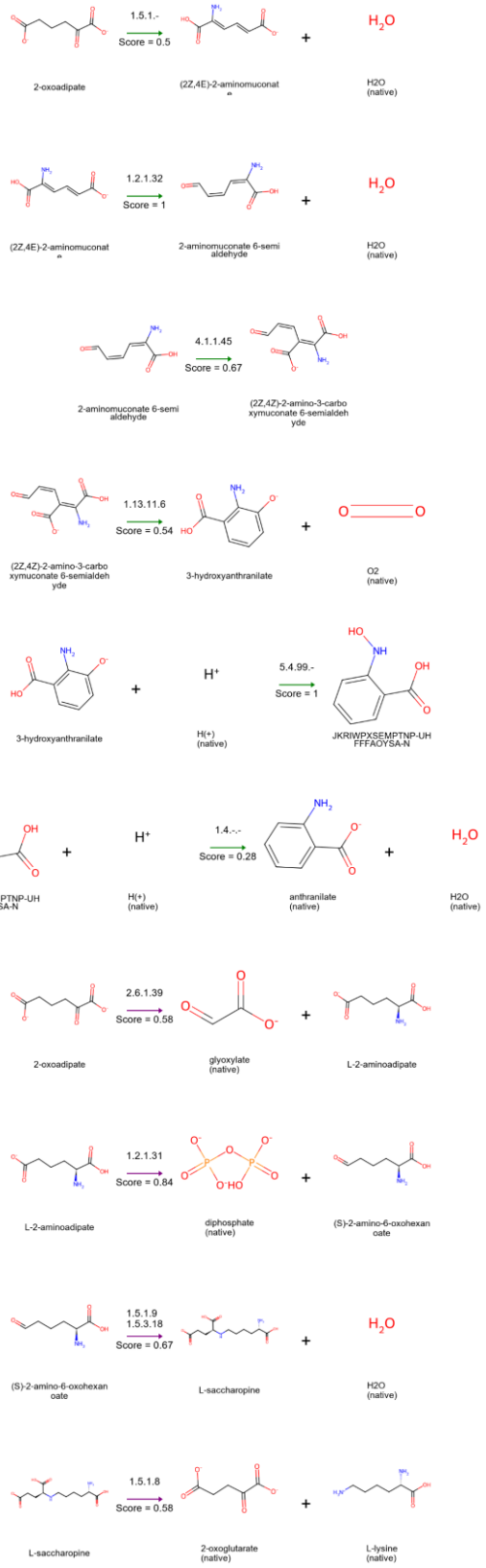




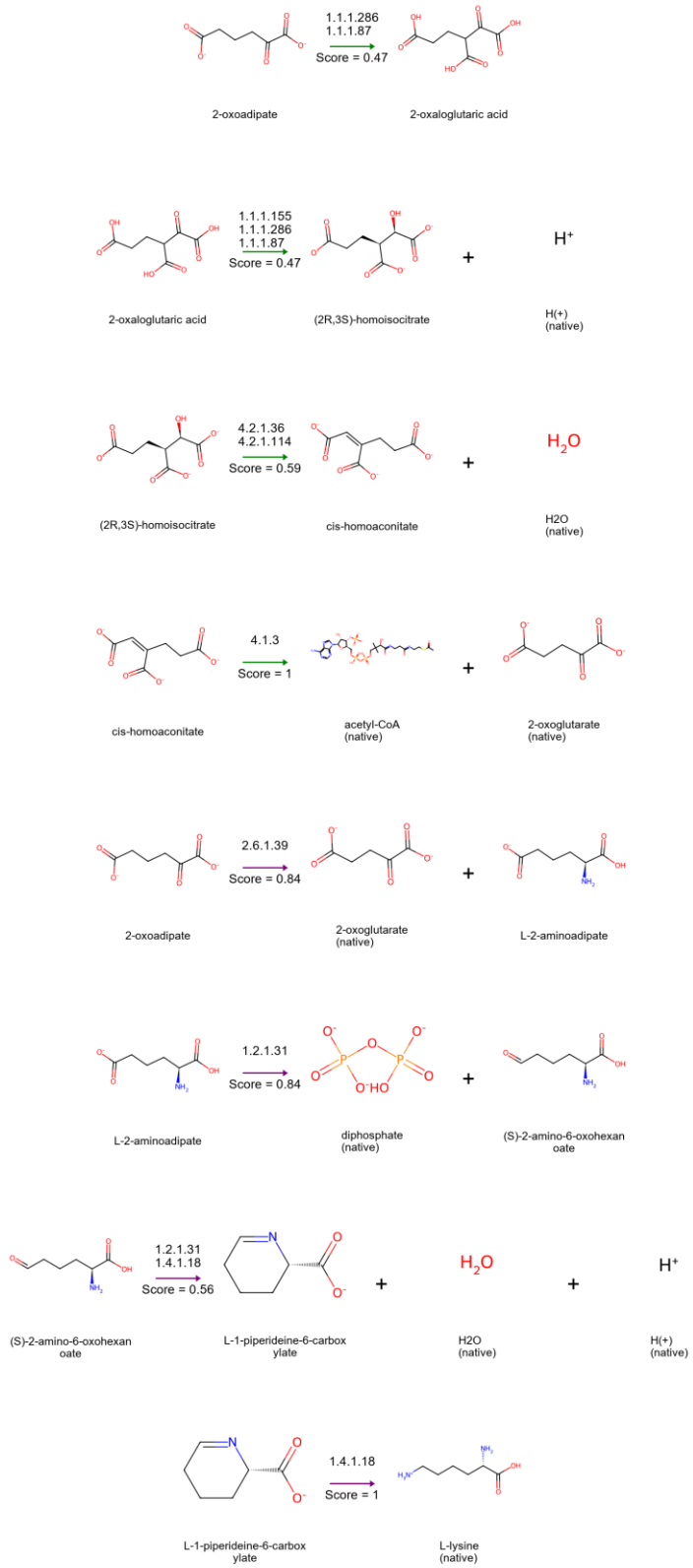




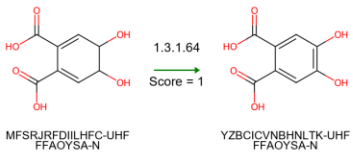
Pathway ID: C237574

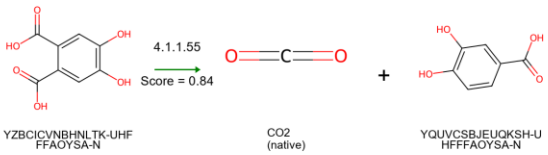


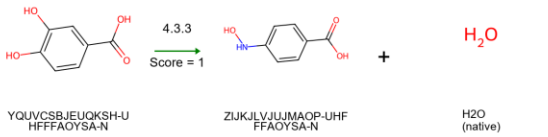
Pathway ID: C238429

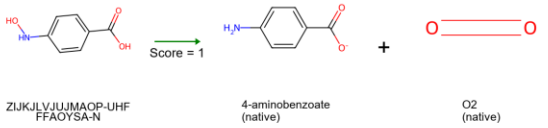


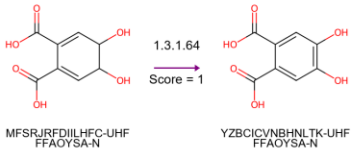
Pathway ID: C239696

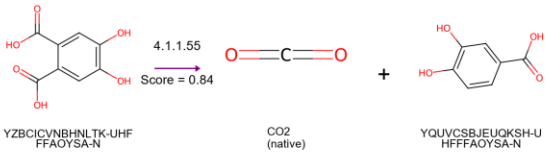

1.3.1.64
Score = 1
MFSRJRFDIILHFC-UHF
FFAOYSA-N
YZBCICVNBHNLTK-UHF
FFAOYSA-N

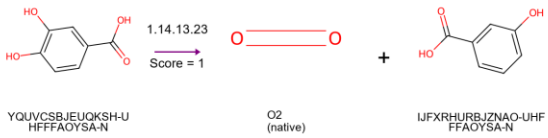

4.1.1.55
Score = 0.84
YZBCICVNBHNLTK-UHF
FFAOYSA-N
CO2
(native)
YQUVCSBJEUQKSH-U
HFFFAOYSA-N

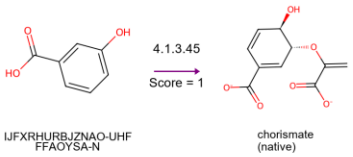

4.3.3
Score = 1
YQUVCSBJEUQKSH-U
HFFFAOYSA-N
ZIJKJLVJUJMAOP-UHF
FFAOYSA-N
H2O
(native)

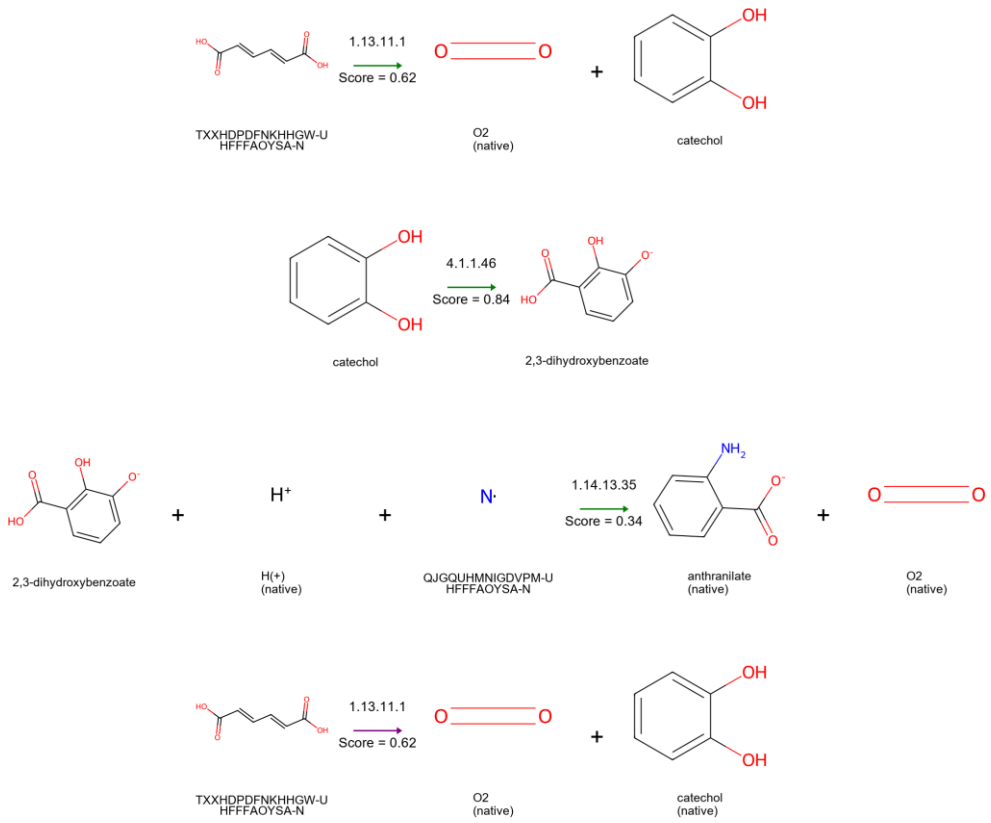

Score = 1
ZIJKJLVJUJMAOP-UHF
FFAOYSA-N
4-aminobenzoate
(native)
O2
(native)


1.3.1.64
Score = 1
MFSRJRFDIILHFC-UHF
FFAOYSA-N
YZBCICVNBHNLTK-UHF
FFAOYSA-N

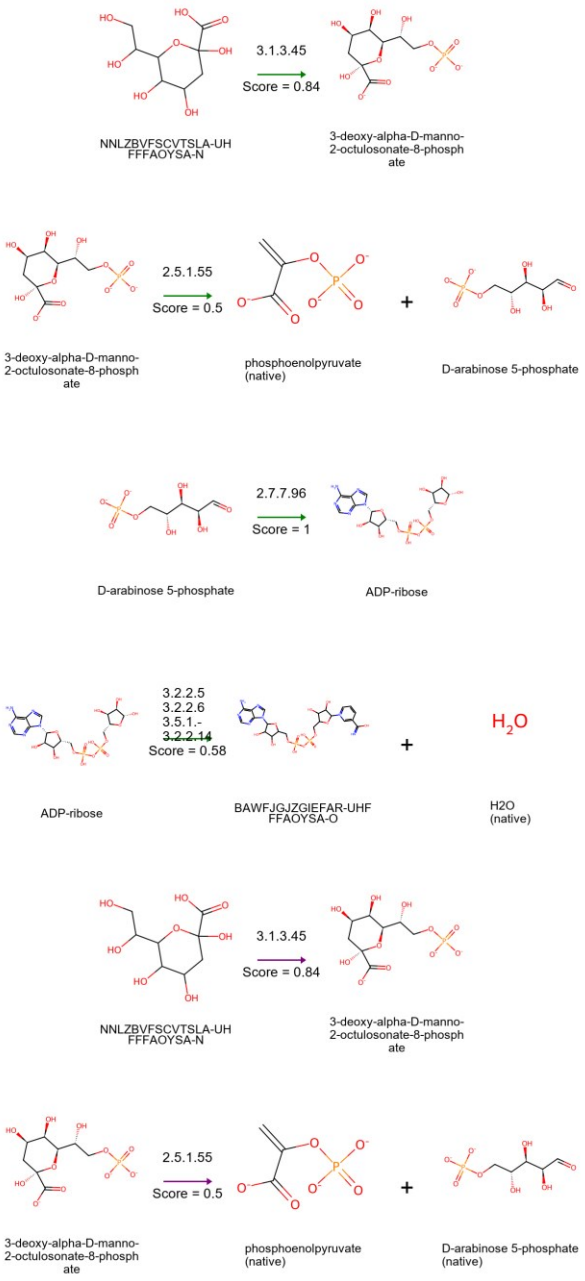

4.1.1.55
Score = 0.84
YZBCICVNBHNLTK-UHF
FFAOYSA-N
CO2
(native)
YQUVCSBJEUQKSH-U
HFFFAOYSA-N


1.14.13.23
Score = 1
YQUVCSBJEUQKSH-U
HFFFAOYSA-N
O2
(native)
IJFXRHURBJZNAO-UHF
FFAOYSA-N

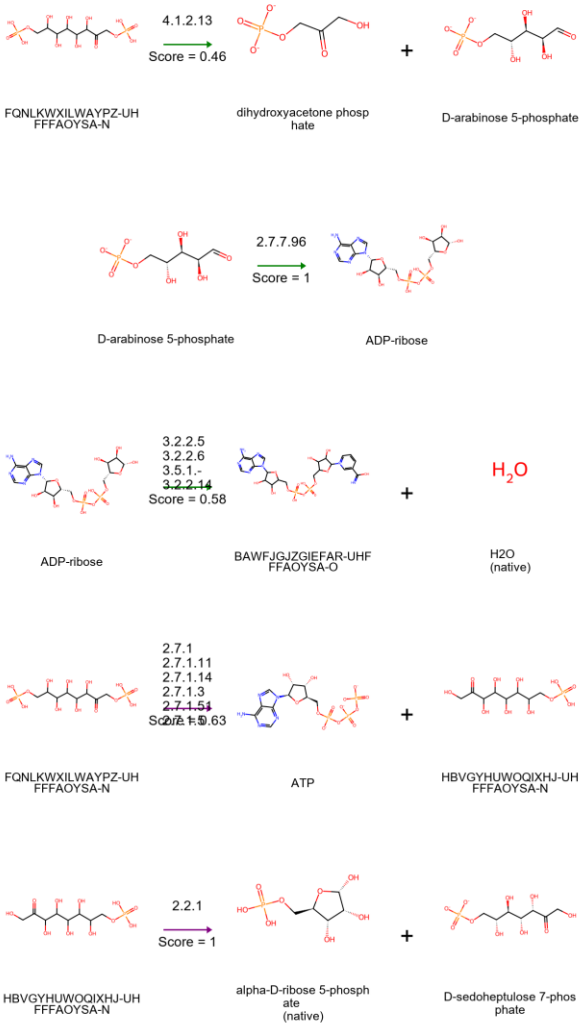

4.1.3.45
Score = 1
IJFXRHURBJZNAO-UHF
FFAOYSA-N
chorismate
(native)



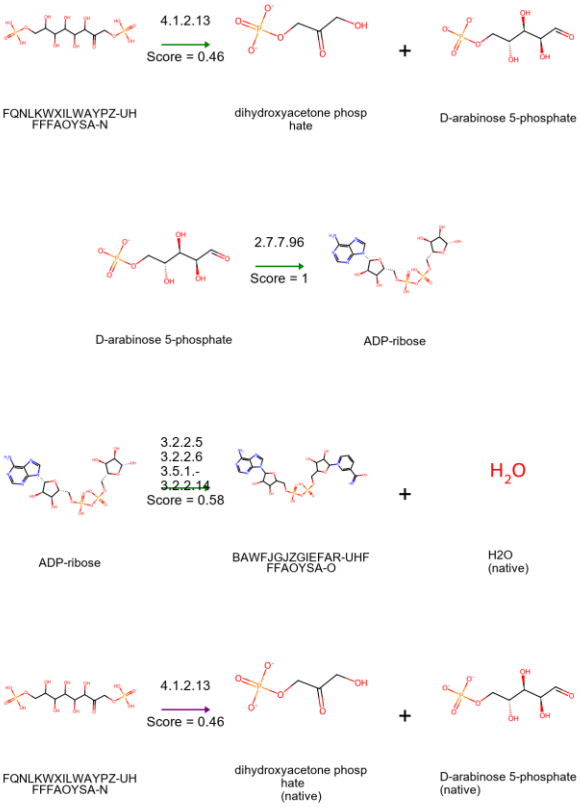
Pathway ID: C279252



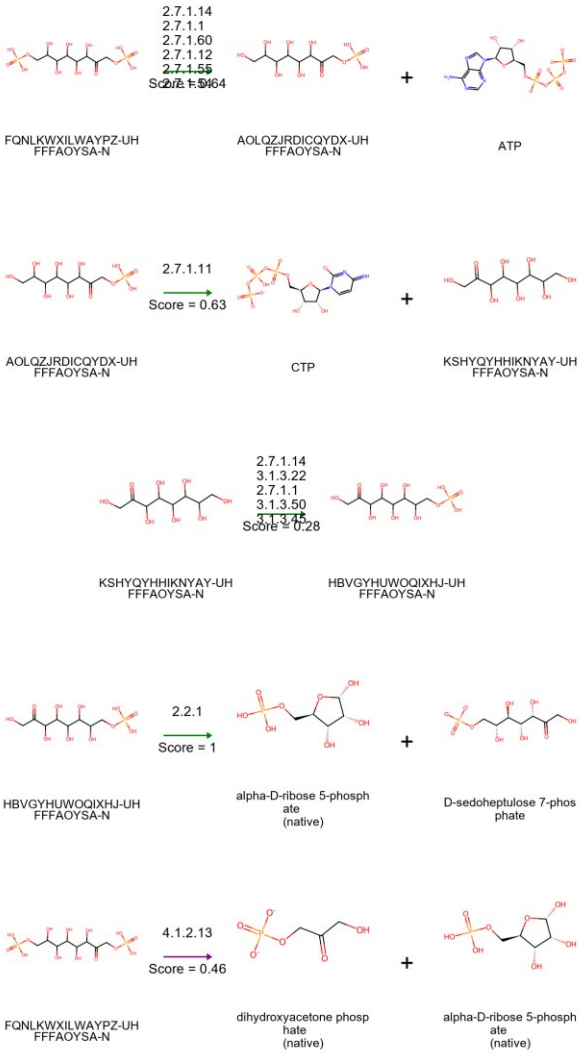
Pathway ID: C280087



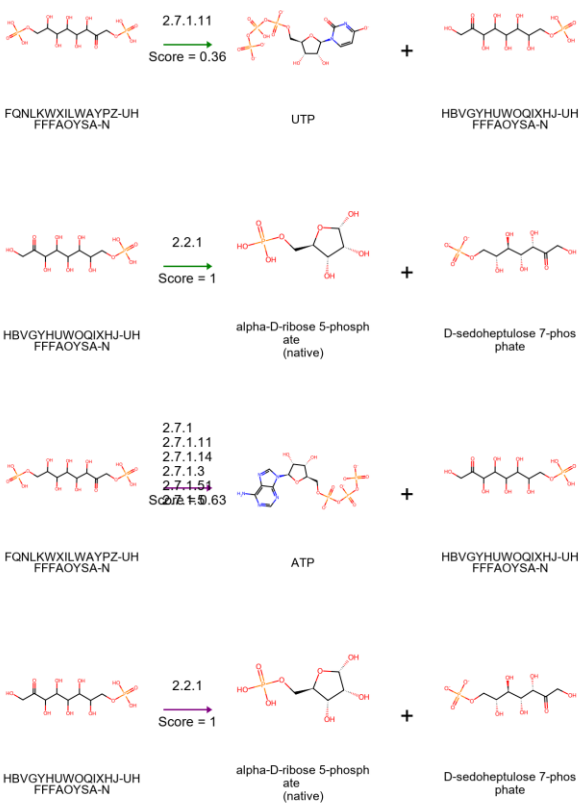
Pathway ID: C39448



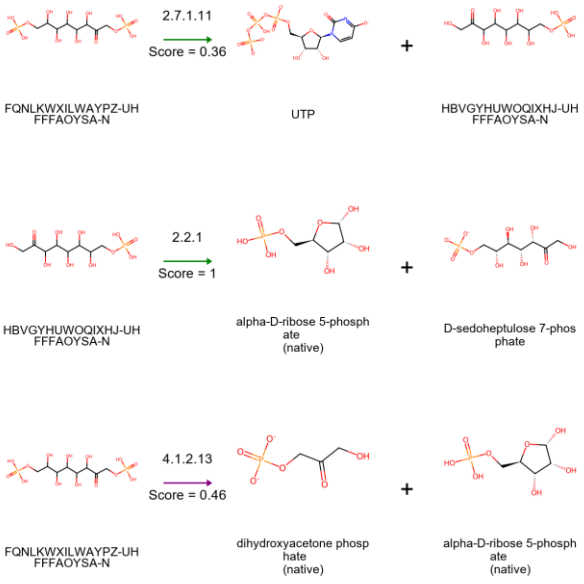
Pathway ID: C39454



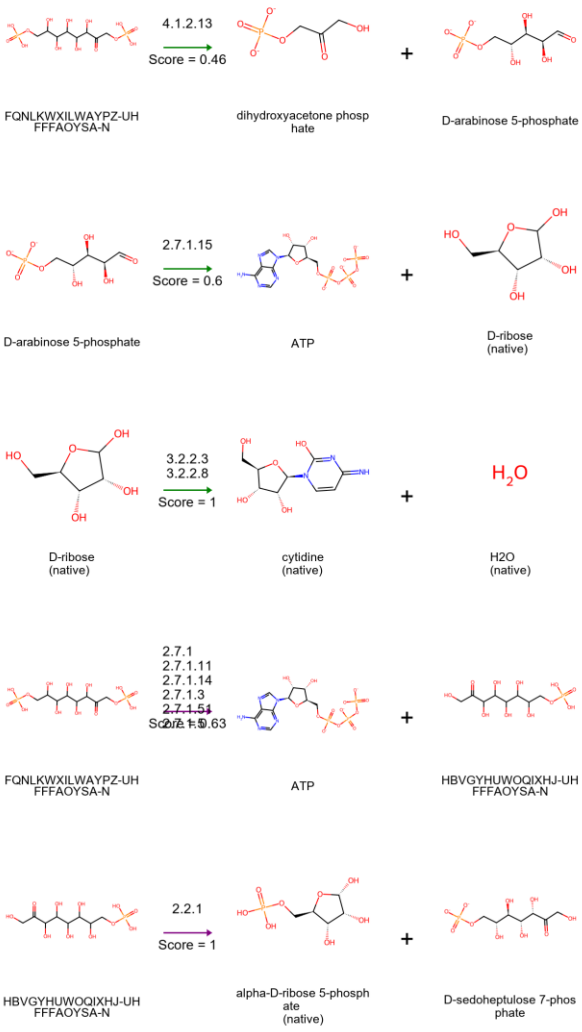
Pathway ID: C39462



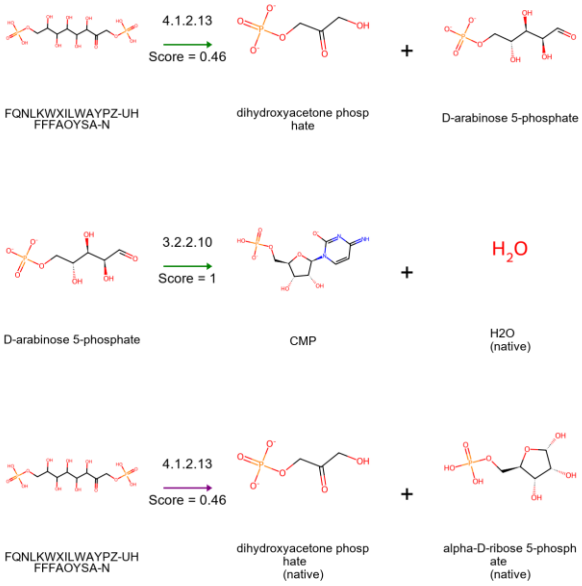
Pathway ID: C39468



Pathway ID: C39472

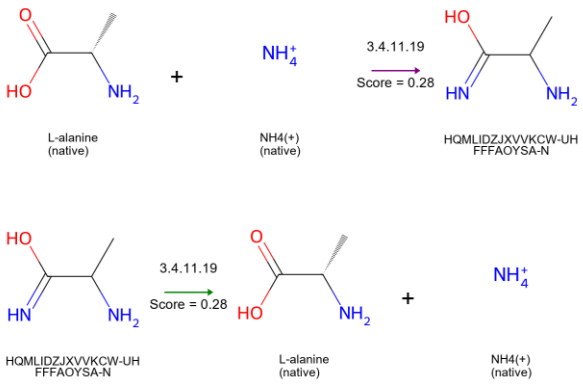


Pathway ID: C39498

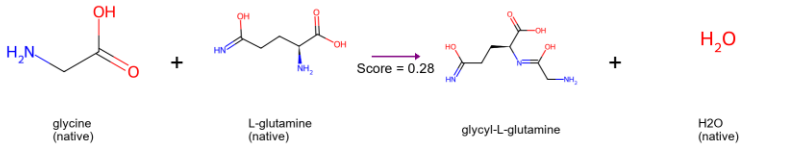


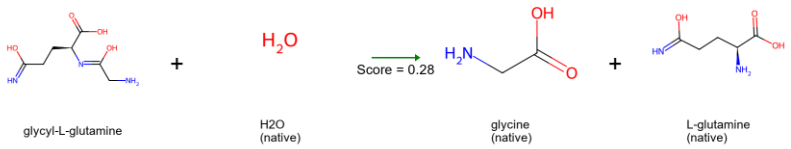
Pathway ID: C39612

9.1.2 Exemplary Nitrogen Exchange Pathways

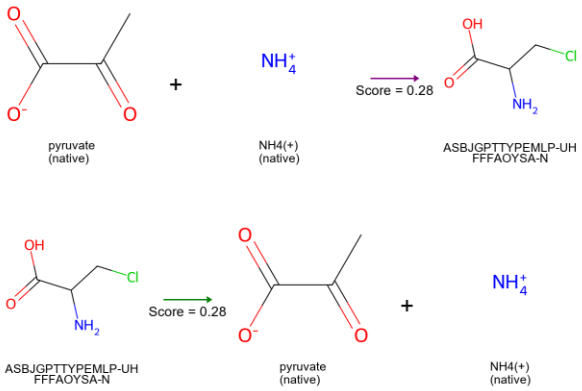


Pathway ID: N96555

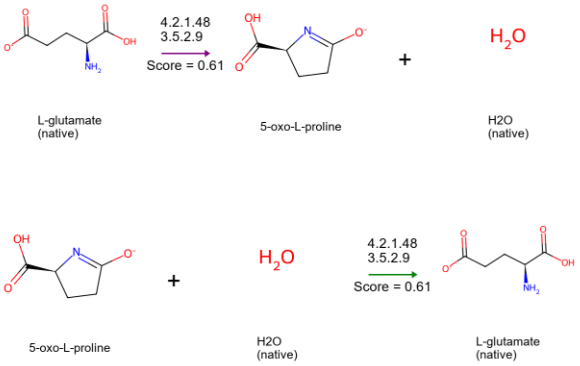




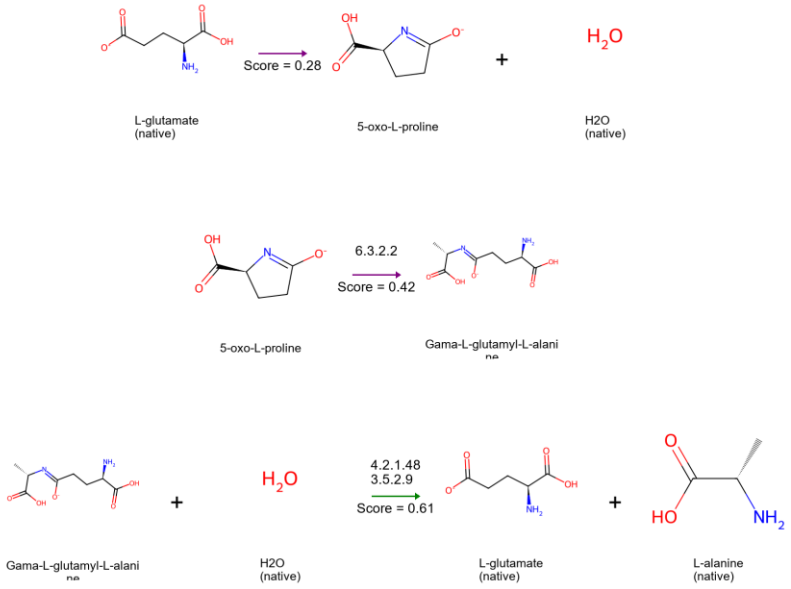
Pathway ID: N99366



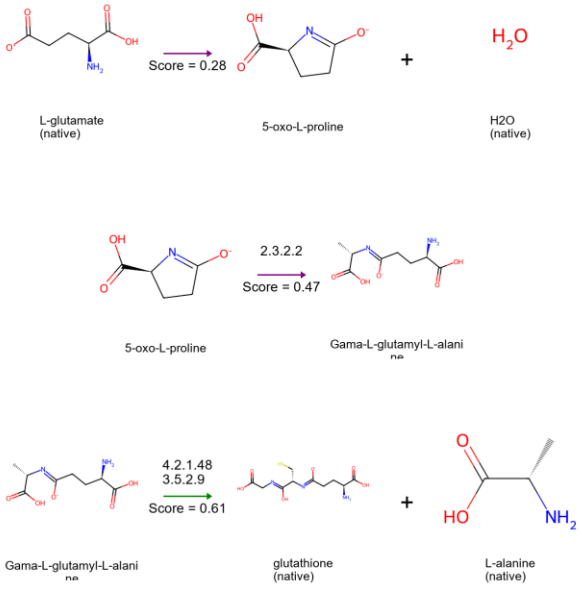
Pathway ID: N99387



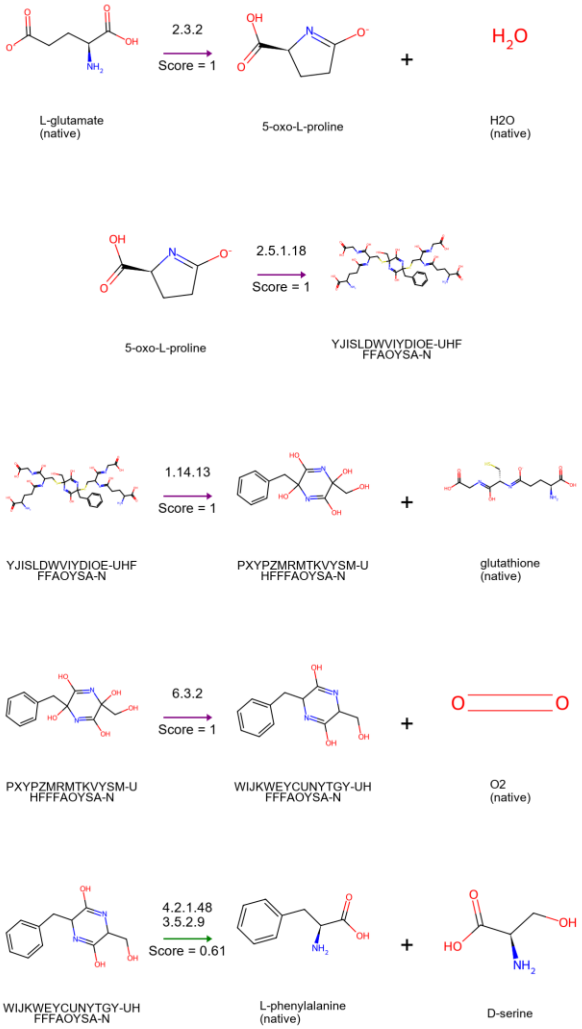
Pathway ID: N103647



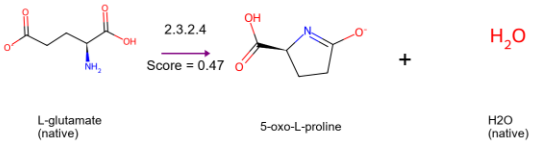
Pathway ID: N103648

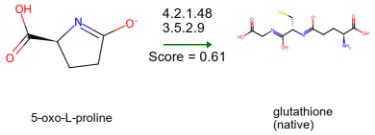


Pathway ID: N103650

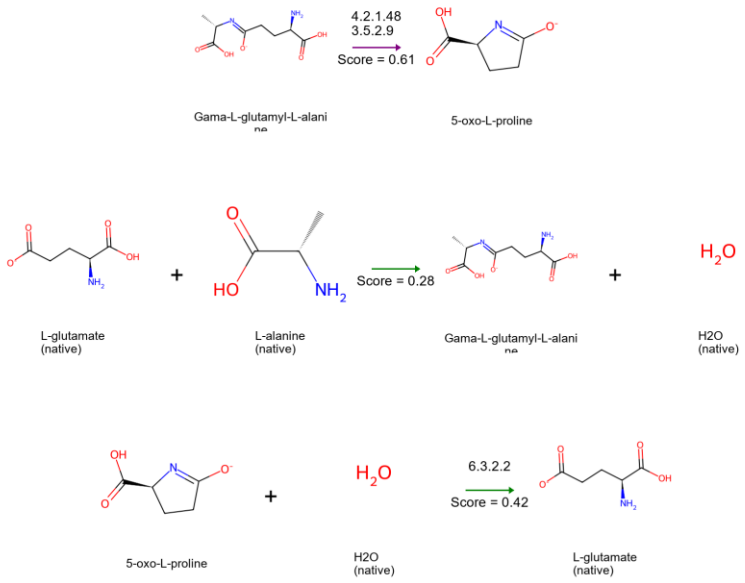


Pathway ID: N103651

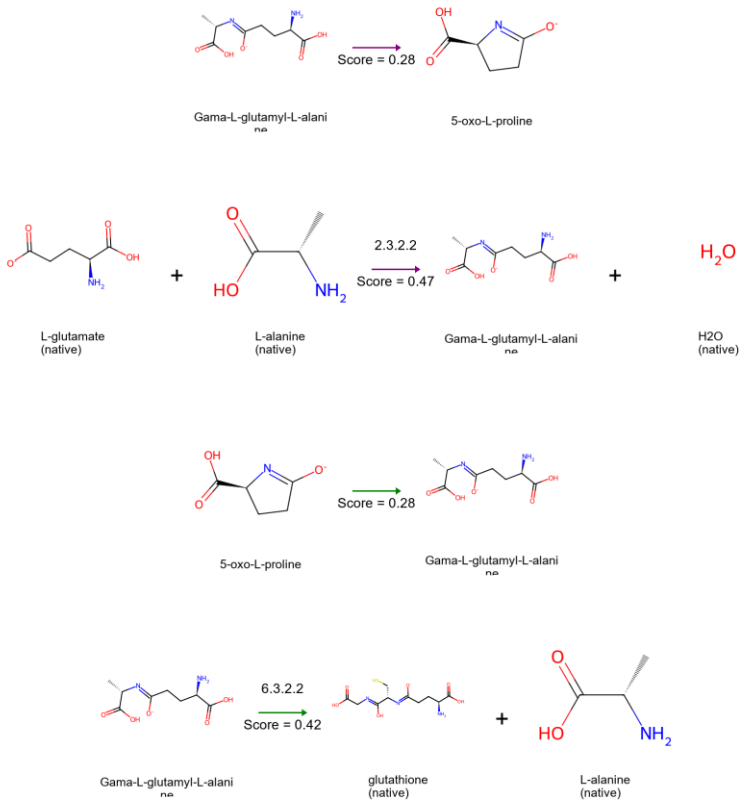




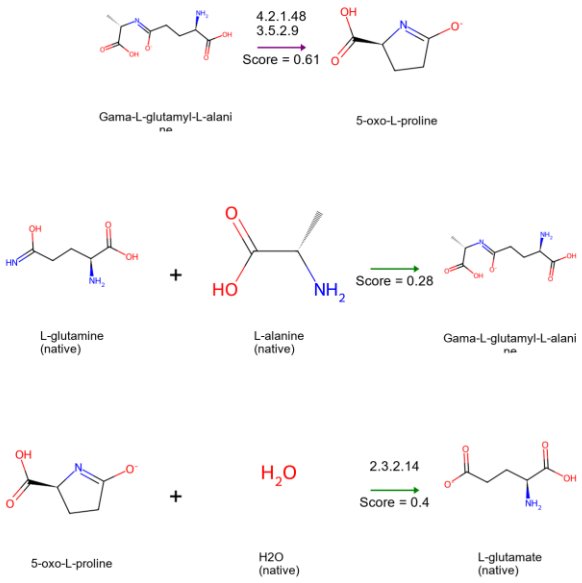
Pathway ID: N103652



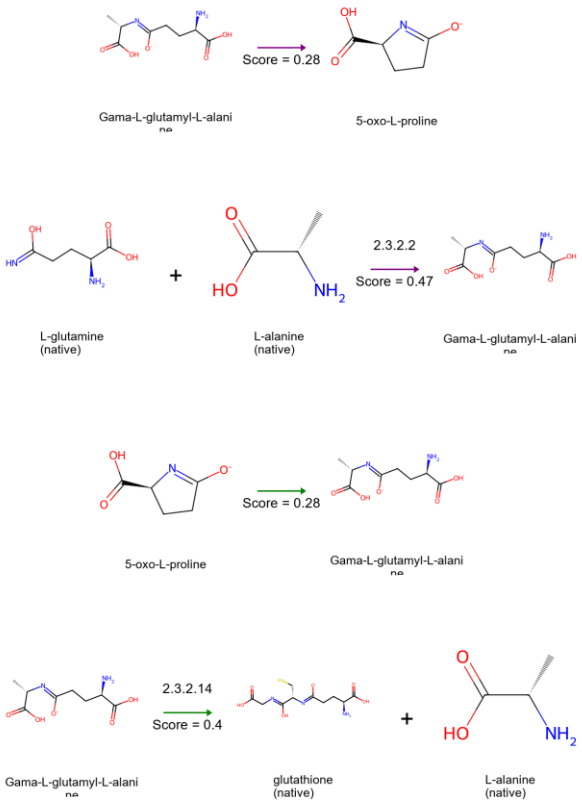
Pathway ID: N103654



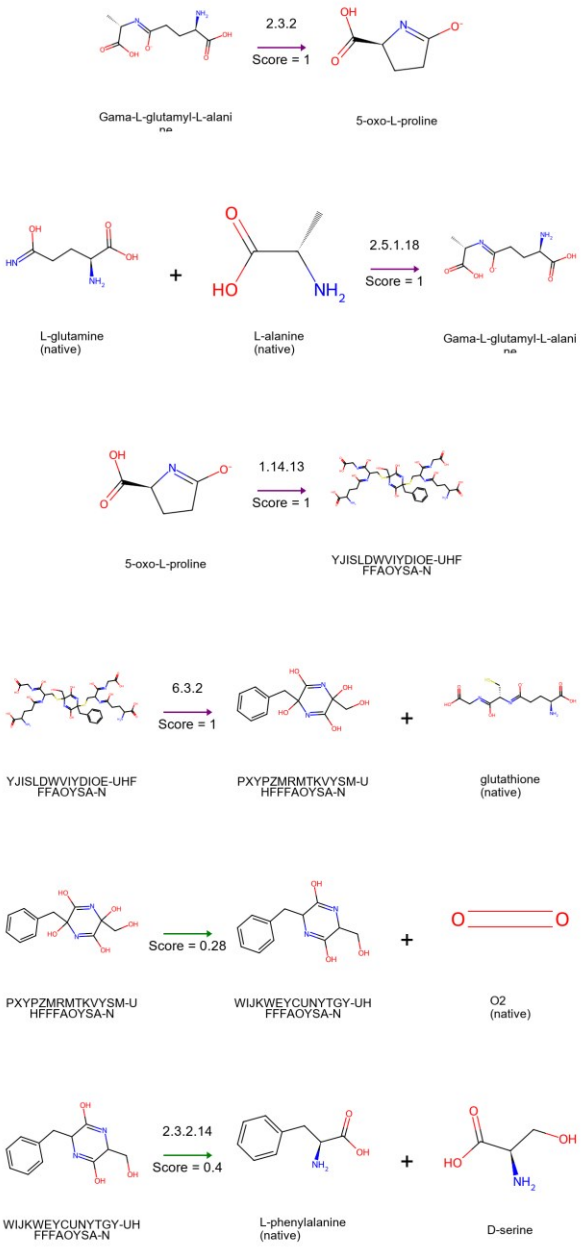
Pathway ID: N103657



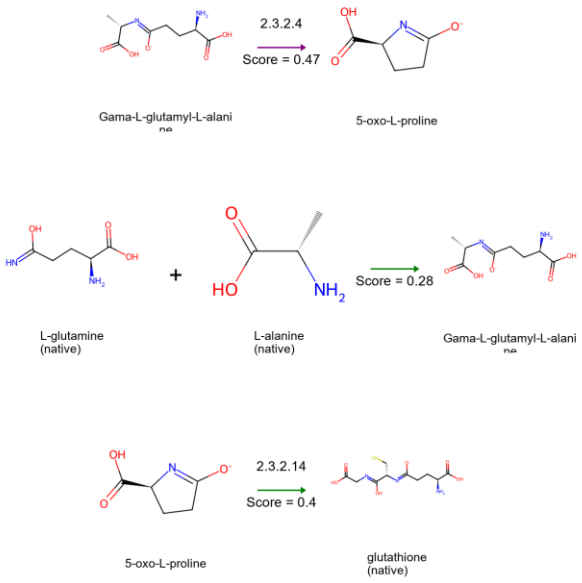
Pathway ID: N103689



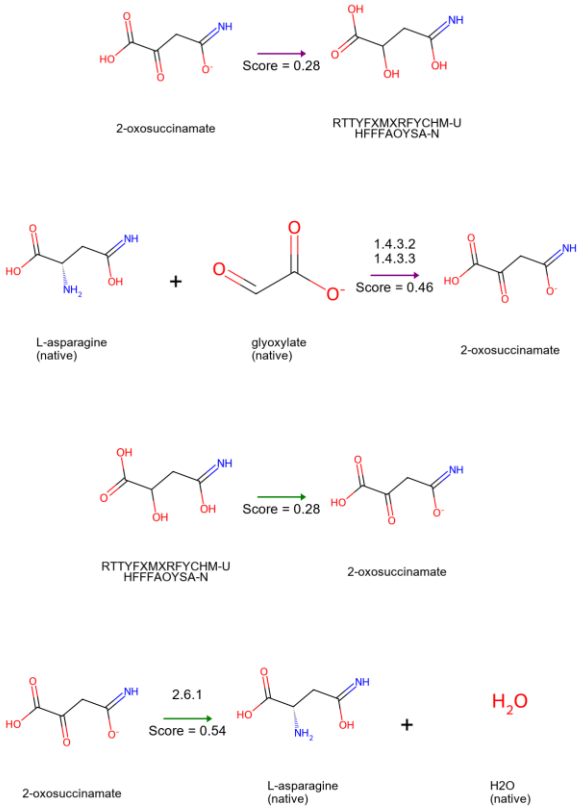
Pathway ID: N103692



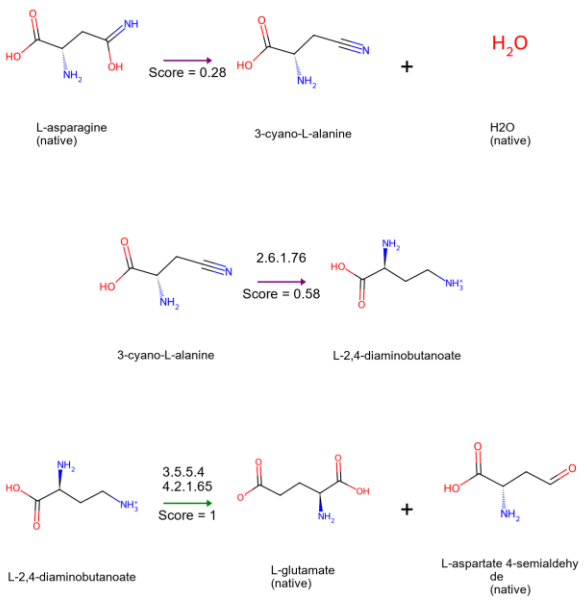
Pathway ID: N103693



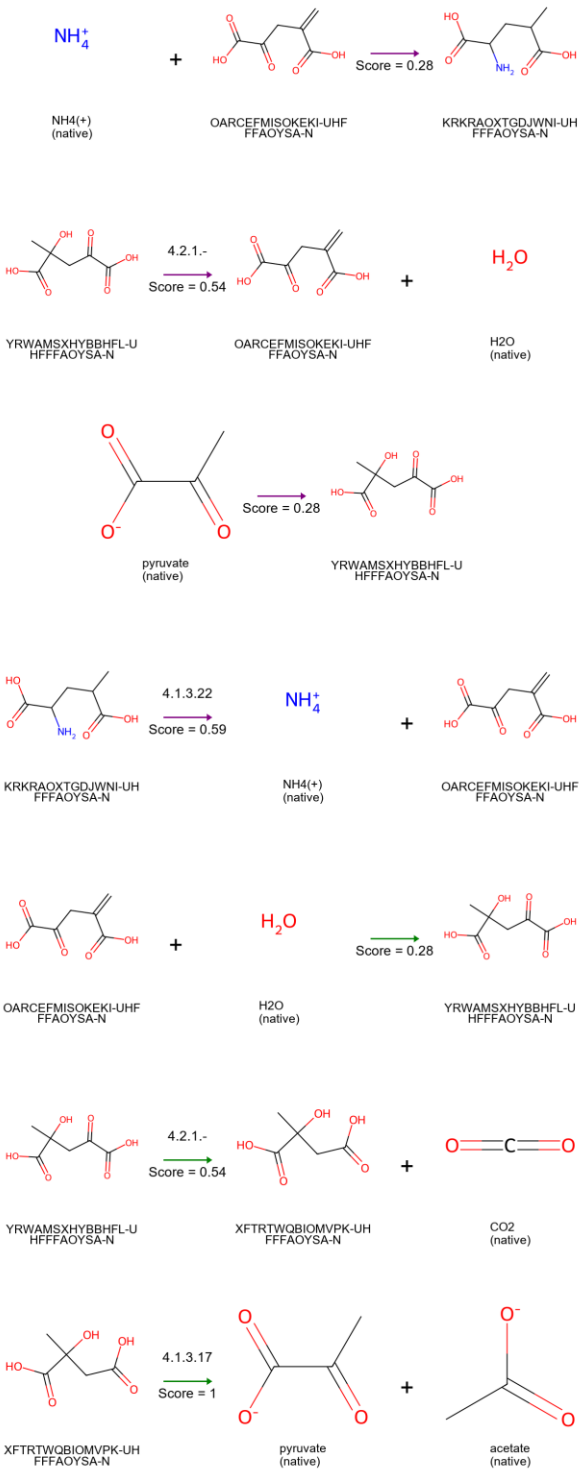
Pathway ID: N103694



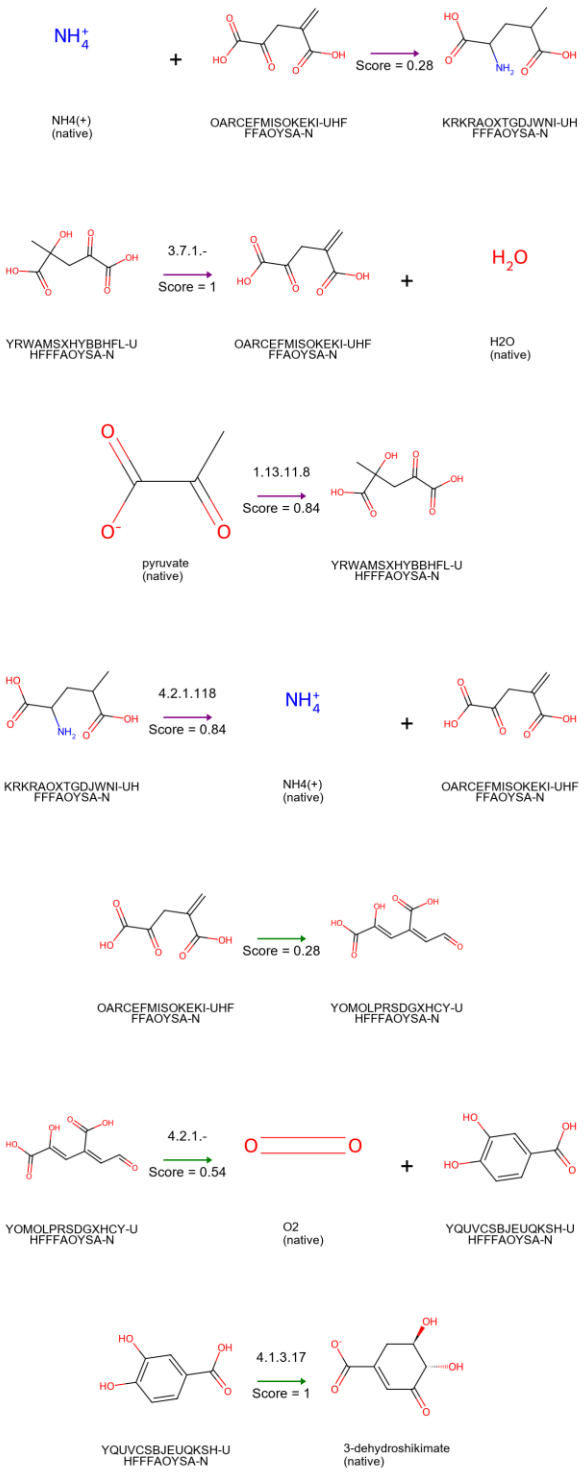
Pathway ID: N109258



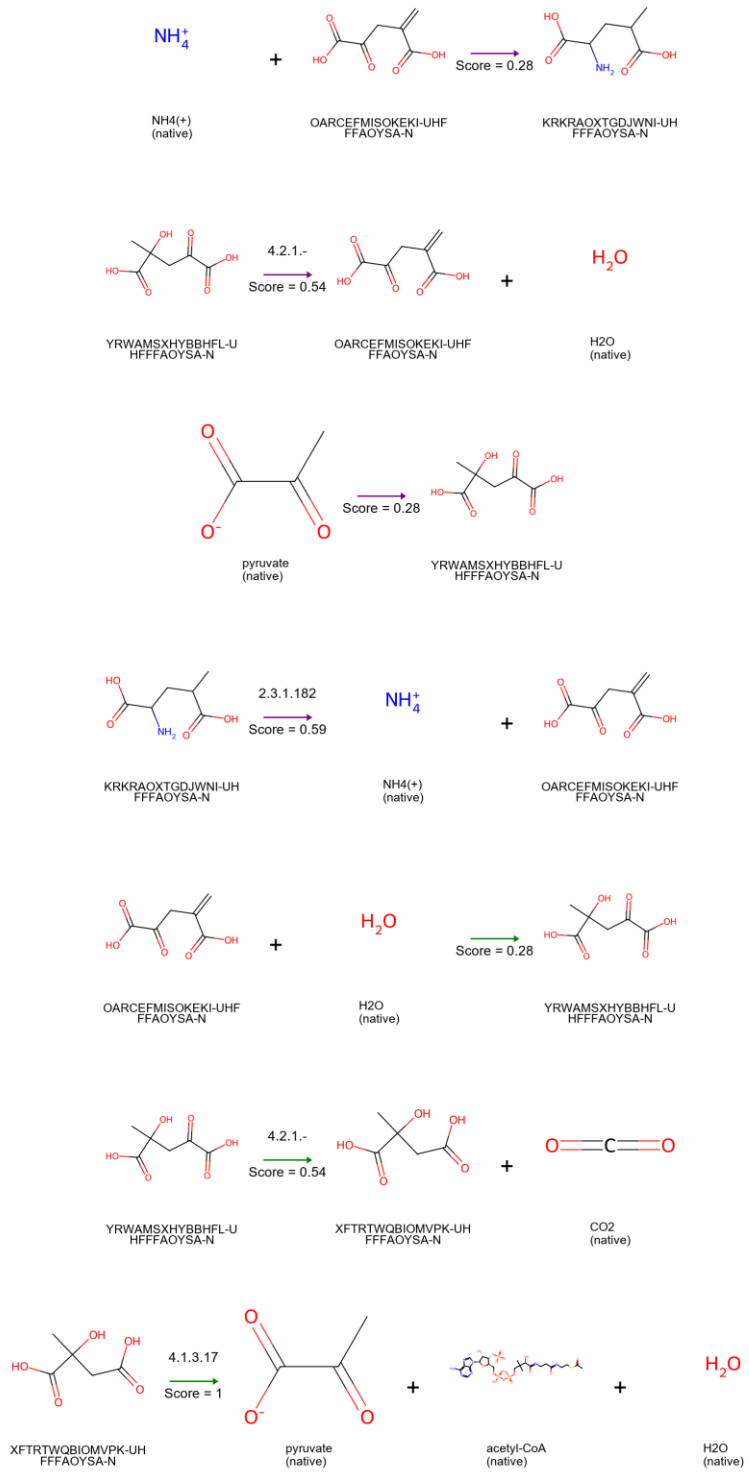
Pathway ID: N117702



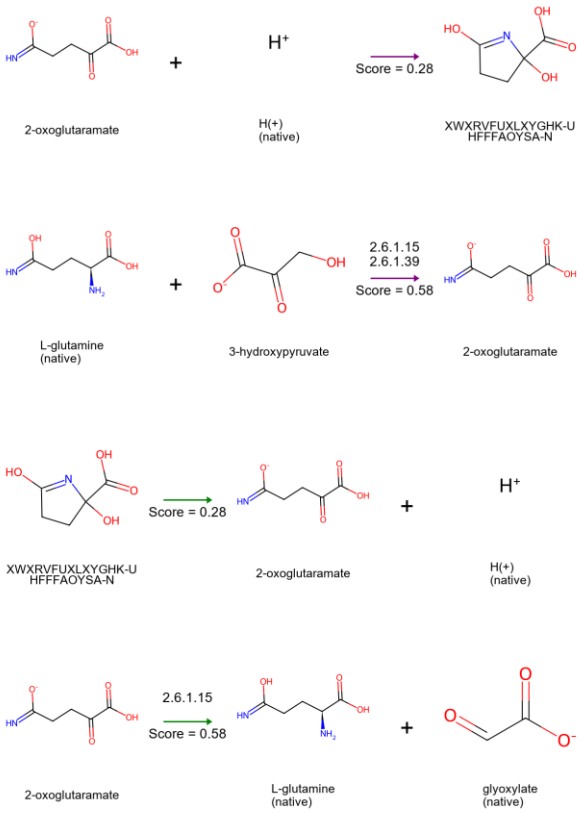
Pathway ID: N128624



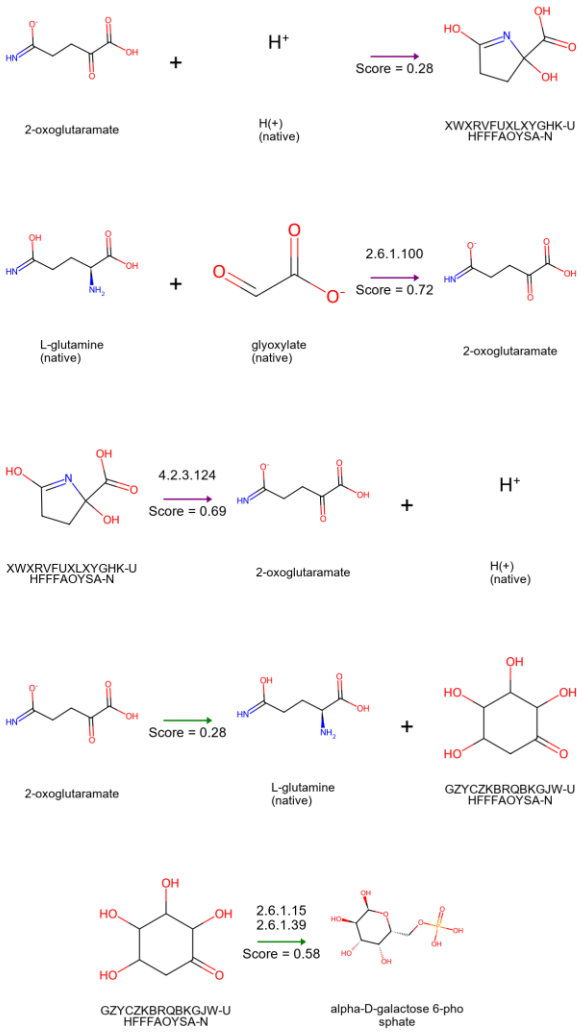
Pathway ID: N128625



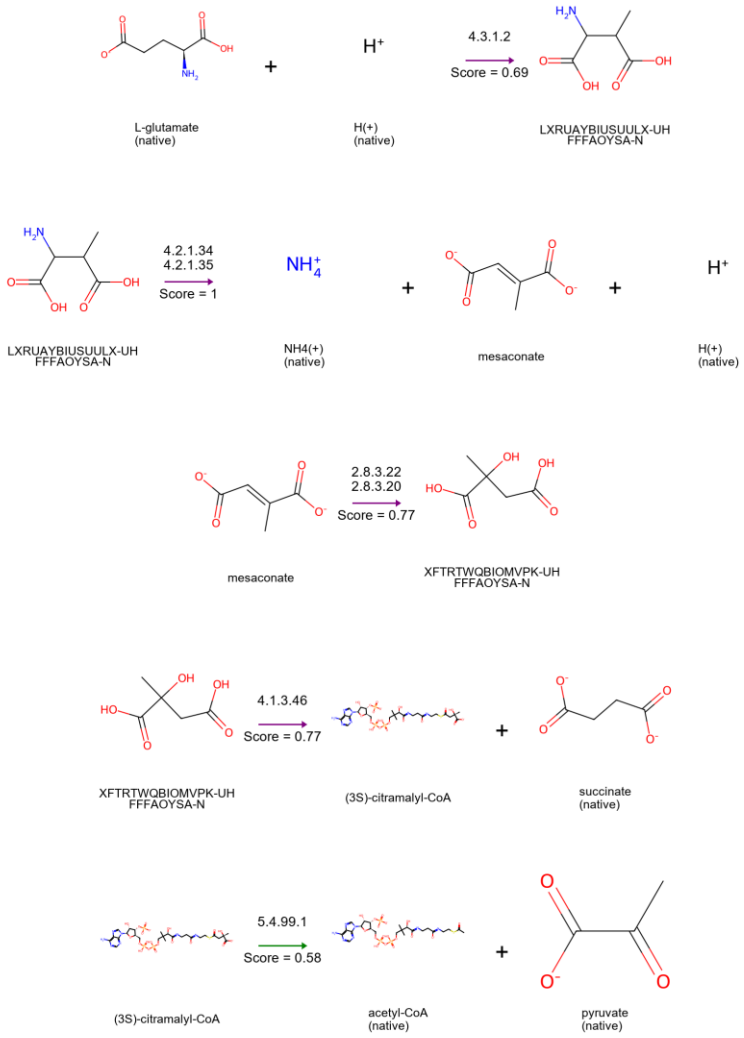
Pathway ID: N128638



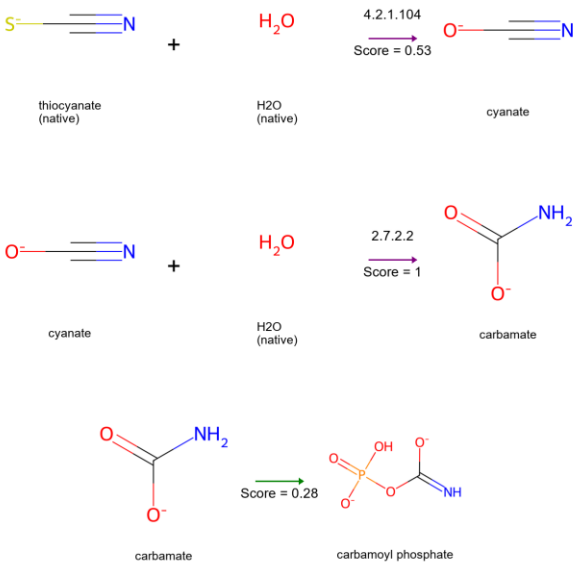
Pathway ID: N49720



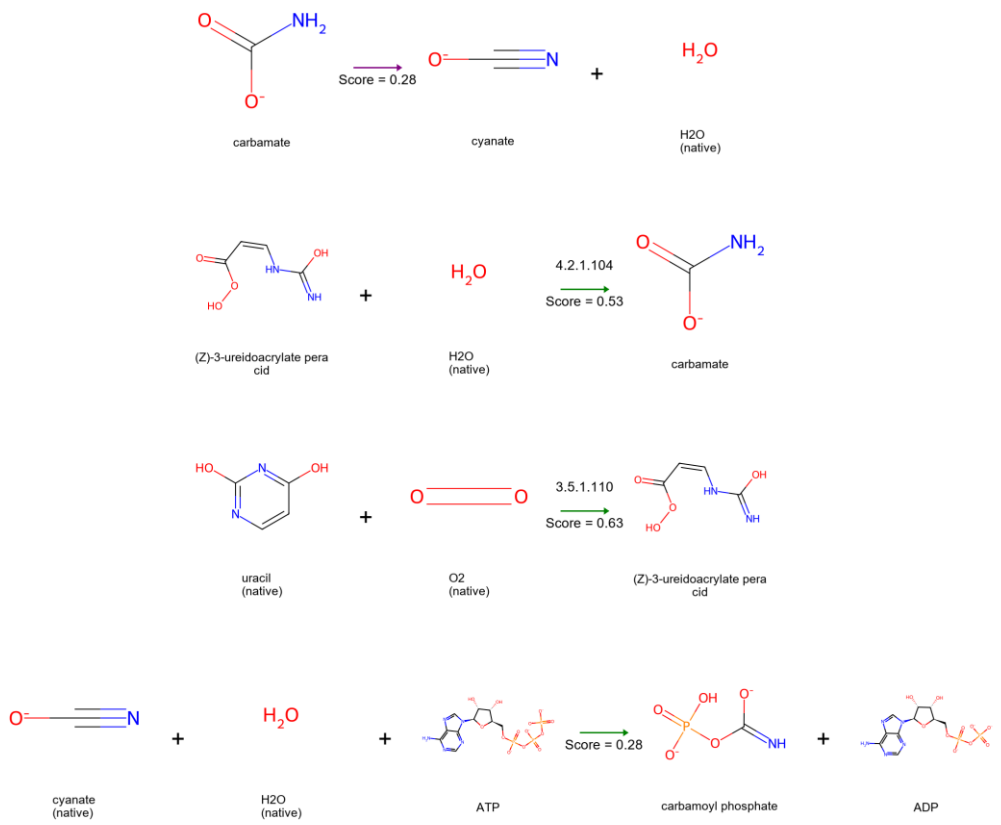
Pathway ID: N49737



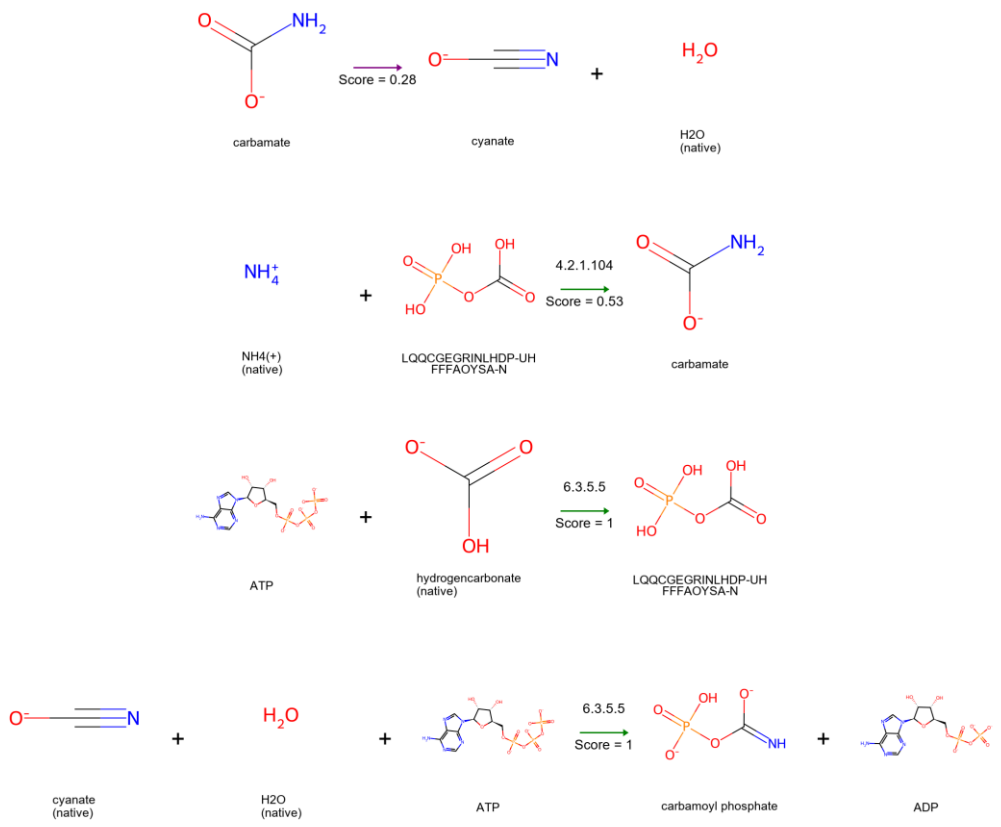
Pathway ID: N78437



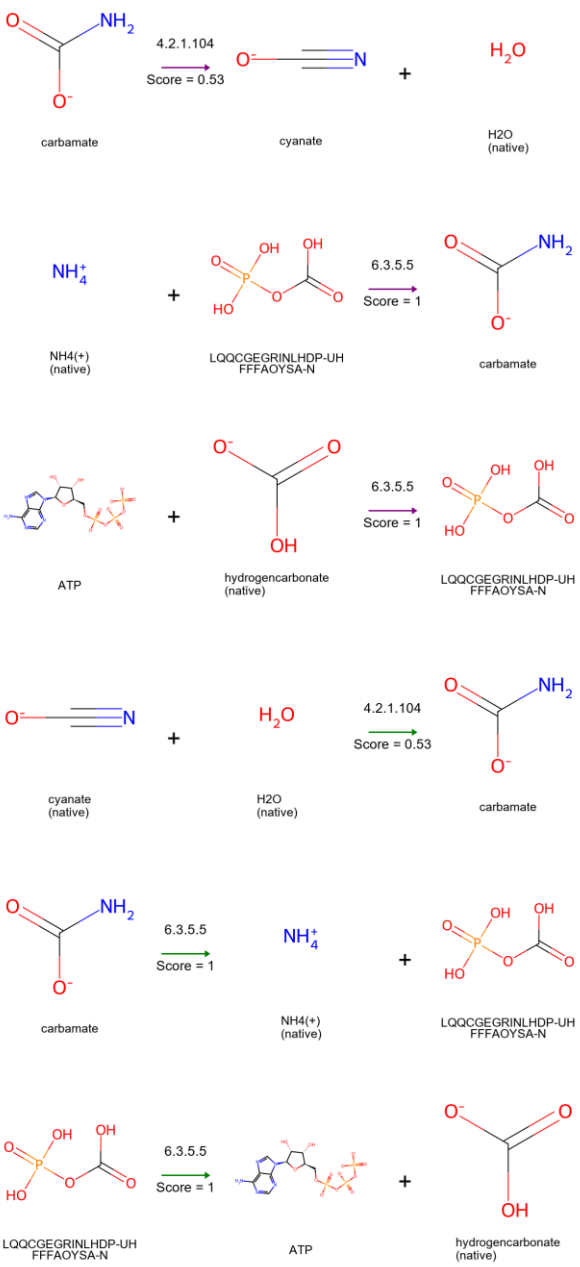
Pathway ID: N91057



Pathway ID: N91088



Pathway ID: N91112



Pathway ID: N91117

9.1.3 DNA sequences

Supplementary table 1: Sequences of DNA parts ordered from IDT. Capitalised sections represent BASIC prefix and suffix sequences.

ID	Length (bp)	Description	Target strain	Sequence
AV_nifL_down_BASIC	187	nifL downstream region	*Azotobacter vinelandii DJ*	TCTGGTGGGTCTCTGTCCacagcgac agggaatgcccatgaatgcaaccatccctcagcgctc ggccaaacagaacccggtcgaactctatgacctgcaa ttgcaggccctggcgagcatcgcccgcacgctcagc cgcgaacaacagatcgacgaactgctcgaacaGG CTCGGGAGACCTATCG
AV_nifL_up_BASIC	187	nifL upstream region	*Azotobacter vinelandii DJ*	TCTGGTGGGTCTCTGTCCaatccatgtt tcgaggtttttccgagcaattcggcgcacccgggcgat taaggtgcggcacaggatttgctaatcttctctcaggcc caacacgcccctccggcggacgcagccgcgctcgc cggttttcttggatagacgaggcacagcGGCTCG GGAGACCTATCG
AV_GOGAT_up_BASIC	187	GOGAT upstream region	*Azotobacter vinelandii DJ*	TCTGGTGGGTCTCTGTCCtgccggcg gaccggaccgactgccggccgtcgcgccagtacatg gctcgttcatcgaggcattgcacggccggcgacgag gcagccaaccgccccggcccgcaagccgtcagccc ggaacctttccgaccagcctcaggaaacccgccgaG GCTCGGGAGACCTATCG
AV_GOGAT_down_BASIC	186	GOGAT downstream region (NOTE: 1bp change to avoid BsaI site)	*Azotobacter vinelandii DJ*	TCTGGTGGGTCTCTGTCCacggcctcg cggaggctagcgccgtttcacccggcgcgggcgtcc ggtcggcgtccgttccgcggcgacgccgccgtagta gTtctcgtgggtggtactggccttggcgtggccgaga tcgacctgcaggttcctgagcggccgatgcggGGC TCGGGAGACCTATCG
ybjL_BASIC	1722	ybjL glutamate transporter from *E. coli*	*Synechocystis 6803*	TCTGGTGGGTCTCTGTCCatgaatataa acgtcgccgaattgttaaatgggaattacattctgttatt atttgtggtcctcgcgcttgggctatgtctcggaaagtt acgacttggttcgatccaactgggtaattccattggcgt tttagtcgtatcgctgttattaggccaacaacatttcagc attaacaccgatgcgcttaatcttggctttatgctgtttatt ttctgcgtcggggtcgaagccggaccgaacttttttcc atttttttttcgcgatgggaaaaattacctaatgttagcact ggtgatggttggcagtgcgctggtgatcgccttagggt taggtaagctgtttggctgggatattggcctgacggcc ggtatgttagcaggctctatgacgtcgacaccggttct ggtcggtgctggcgatacactgcgtcattccggcatg gaaagcaggcagctctcactggcactggataatctga gcctcgggtatgccttaacctatttaatcggtctggtga gtttgattgttggtgcgcgttacttgccgaaattgcagc atcaggacttacagaccagcgcccagcaaatcgccc gcgaacgtggcctggacactgatgccaaccgtaagg tttatttaccggtgatccgcgcctatcgcgtcggcccg gaactggtggcctggaccgacggcaaaaatctgcgt gaactgggtatttatcgacaaaccggctgctacattga acgtattcgacgtaacgggattctggcaaatccagac ggtgatgccgtgctacaaatgggcgatgaaatagcgt tggtaggctatcccgacgcccatgcccgactcgatcc cagcttccgtaacggtaaagaagttttcgatcgtgacct tctcgacatgcgtatcgtcactgaagaagtggtcgttaa aaaccataacgctgtaggtaaacgtctcgcacaactg aagttgaccgatcacggttgcttccttaaccgcgtcatt cgtagccagattgagatgccgatagatgacaacgtcg tgcttaacaaaggtgacgttttacaagtcagcggcgat gcccgccgcgtaaaaaccatcgccgatcgcatcggc tttatctcgattcacagccaggtcactgacctgctggca ttctgcgccttctttgttattgggctgatgatcgggatgat caccttccagttcagcacattcagtttcggcatgggga acgctgccgggttgttattcgccggaattatgctgggct ttatgcgtgctaaccacccgaccttcggttacattccgc agggtgcattaagcatggtgaaagagttcggcttgatg gtgtttatggcaggcgttggtctgagcgccggtagcg

				gtattaataacggcctgggcgcgattggcggtcagat gttgattgccggattgattgtcagtctggtgcccgtggtt atctgtttcttgttccggtgcttatgtattgcgaatgaaccg cgcgctgttgttcggcgcaatgatgggcgcacgtacc tgcgcgccggcaatggagatcatcagtgatacagctc gcagtaacatcccggcgctgggctatgcgggcacct atgcaatcgccaacgtcctgctgacgctggcagggac aatcatcgtcatggtatggccaggattaggataaGG CTCGGGAGACCTATCG
cscB_BASIC	1284	*cscB* sucrose-proton symporter	*Synechocystis 6803*	TCTGGTGGGTCTCTGTCCatggccctc aatatccccttccgtaacgcatattatcgctttgccagct cctattcttttctgtttttcattagctggagcctatggtggt ccctctacgcgatttggttaaaaggtcatttaggcctca ccggcacggagttgggtacattatactctgtcaatcaat tcacttctattttgttcatgatgttctatggtatcgtgcagg ataagctagggctaaaaaagcctttgatctggtgcatgt cctttattctagtgctcaccgggccgttcatgatctacgt ttatgaaccgttattgcaaagtaacttctctgtggggctc attttgggtgcattatttttttgggctaggttatctggccgg atgtggtttactggacagtttcaccgaaaaaatggcgc gtaatttccacttcgagtatggaacggcgcgtgcgtgg gggagtttcgggtatgctatcggcgccttttttgctggta ttttttttcagtatttctccccacatcaatttttggctcgtctc cttgtttggtgcagtctttatgatgattaacatgtgcttca aggataaggatcatcaatgtgtagccgctgacgcgggg tggggtgaaaaaggaggatttcatcgcagtattcaaag accggaatttctgggtatttgtgattttcattgttgggacg tggtccttttacaacatctttgatcagcagttatttccggtt ttttatgcagggctattcgagtcccatgatgtcgggact cgtttatacggatatctaaactccttccaagtagtattgg aagcgctgtgtatggccatcatcccgttctttgtcaatc gagtaggccccaagaatgctttactcatcggggtggtt attatggctctgcgaatcctgtcttgcgccctcttcgtga acccgtggatcattagcttagtaaaactactacacgcg atcgaggtcccactctgcgtgatcagcgtctttaaatat agtgtggctaatttcgacaaacgtctaagcagcaccat ctttctaattggttttcagattgcctcctctttaggcattgt actcttatctacccctactggcattttgtttgatcatgctg gttatcagaccgtgttttttgcgattagtggtattgtatgt ctaatgttattgttcggtatcttctttctaagtaaaaagcgt gagcagattgtgatggaaacacccgttccatccgcaat ttagGGCTCGGGAGACCTATCG

Supplementary table 2: Sequences of BASIC linkers used in this study

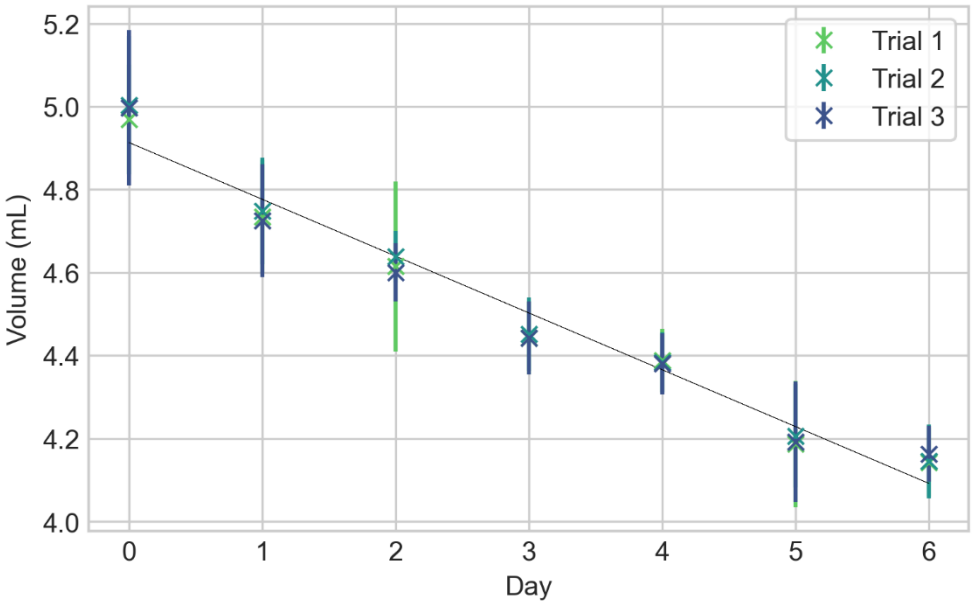
ID	**Sequence**
Neutral linker L1	GGCTCGTTACTTACGACACTCCGAGACAGTCAGAGGGTATTTATTGAACTAGTCC
Neutral linker L2	GGCTCGATCGGTGTGAAAAGTCAGTATCCAGTCGTGTAGTTCTTATTACCTGTCC
Neutral linker L3	GGCTCGATCACGGCACTACACTCGTTGCTTTATCGGTATTGTTATTACAGAGTCC
RBS linker RBS1	GGCTCGTTGAACACCGTCTCAGGTAAGTATCAGTTGTAAATCACACAGGACTAGTCC
Methylated linker LMP	GGCTCGGGTAAGAACTCGCACTTCGTGGAAACACTATTATCTGGTGGGTCTCTGTCC
Methylated linker LMS	GGCTCGGGAGACCTATCGGTAATAACAGTCCAATCTGGTGTAACTTCGGAATCGTCC

Supplementary table 3: List of primers	
Name	**Sequence**
F_amtB_A_vinelandii	gccggagtgatgatggcgaagg
R_amtB_A_vinelandii	tccctcttcctgcccgtcatgg
F_NifL_A_vinelandii_outer	gatttgctaatcttctctcaggc
R_NifL_A_vinelandii_outer	gcaggtcatagagttcgacc
F1_NifL_A_vinelandii	atcttctctcaggcccaacac
F2_NifL_A_vinelandii_inner_forseq	aacgctacctgctgctg
F_UreABC_A_vinelandi_outer	GCCGGCGATGATATCCG
F_UreABC_A_vinelandii_Inner	ACAGGCAGTTGTCGATGG
R_UreABC_A_vinelandii_outer	TCGATTTATGGCGCCTGC
AV_GDH_downstream_F_BASIC_prefix	TCTGGTGGGTCTCTGTCCagcggccccgc
AV_GDH_downstream_R_BASIC_suffix	CGATAGGTCTCCCGAGCCgagttgggttatctgcagt
AV_GDH_upstream_F_BASIC_prefix	TCTGGTGGGTCTCTGTCCggcgtttccagctcc
AV_GDH_upstream_R_BASIC_suffix	CGATAGGTCTCCCGAGCCggtcgtgctcccgaa
AV_glnA_downstream_F_BASIC_prefix	TCTGGTGGGTCTCTGTCCtccagccggcgatac
AV_glnA_downstream_R_BASIC_suffix	CGATAGGTCTCCCGAGCCgatttccgcgaaagacgt
AV_glnA_upstream_F_BASIC_prefix	TCTGGTGGGTCTCTGTCCgggattatagcggatcttgttc
AV_glnA_upstream_R_BASIC_suffix	CGATAGGTCTCCCGAGCCgcgacttcgacatattgtcc
R_scrY-1_BASIC_suffix	CGATAGGTCTCCCGAGCCcaaattgaacttgcaatgtgcatgg
F_scrY-1_BASIC_prefix	TCTGGTGGGTCTCTGTCCcctactacaagaacaagaggtttaccc
R_Avinelnadii_srt_BASIC_suffix	CGATAGGTCTCCCGAGCCgacgcggattgcttcgt
F_Avinelnadii_scrT_BASIC_prefix	TCTGGTGGGTCTCTGTCCcctcaaataaaaacaagagggttattcc
F_K12_ybjL_BASIC_prefix	TCTGGTGGGTCTCTGTCCctagtaaacaagaagttagcctcc
R_K12_ybjL_BASIC_suffix	CGATAGGTCTCCCGAGCCcattttcagggcaacttcagtt
F_AV_scrY_SDM_BsaI_removal	gtgtcgaggtgcgAgtc
R_AV_scrY_SDM_BsaI_removal	gcagtccctggagacTc

F_OEPCR_BASIC_into_pJET	CCGGATGGCTCGAGTTTTTCAGCAAGATTCTGGTGGGTCTCTGTCC
R_OEPCR_BASIC_into_pJET	GGCAGCTGAGAATATTGTAGGAGATCTTCTAGAAAGATCGATAGGTCTCCCGAGCC
pJET_F	AATACGACTCACTATAGGGAGAGC
pJET_R	AAGAACATCGATTTTCCATGGCAG
F_JM01_seq	GGCGTATCACGAGGCCCTTTCG
R_JM01_seq	AGTCTTTCGACTGAGCCTTTCGT
F_JM02_seq	GGCGTATCACGAGGCCCTTTCG
R_JM02_seq	AGTCTTTCGACTGAGCCTTTCGT
seq gifA F (Galmozzi)	gagttacttcctctaataacaacc
seq gifA R (Galmozzi)	ctcttgcagtgttctgttgctgg
gifA mut inner 1 (Galmozzi)	GATGCGCTAATATTAGCAAGTGAAGAATCG
gifA mut inner 2 (Galmozzi)	CGATTCTTCACTTGCTAATATTAGCGCATC
gifA BASIC pre F	TCTGGTGGGTCTCTGTCCgagttacttcctctaataacaacc
gifA BASIC suf R	CGATAGGTCTCCCGAGCCctcttgcagtgttctgttgctgg

9.1.4 Evaporation assay

A simple evaporation assay was set up to determine the amount of liquid lost to evaporation during cultivation in 6-well plates under cultivation conditions used throughout this study (Supplementary Figure 1).



Supplementary Figure 1: Volume of water in each well under cyanobacterial cultivation conditions was shown to decrease by ca. 137 µL daily (Linear regression $R^2 = 0.97$)

9.2 Supplementary Methods

9.2.1 Evaporation assay

A set of seven 6-well plates was filled with sterile MiliQ water, using 5 mL in each well and sealed with microporous tape. The plate was kept at conditions equivalent to culture conditions – 160 rpm at 30°C, 1% CO_2 and 60 µE in a PSI Algaetron AG 230 growth chamber. Once a day, one well plate was collected for measurement, and the volume in each well was measured using a measuring cylinder. The process was carried out in triplicate.

9.2.2 Media Preparation

9.2.2.1 Burke's Medium (B and BN)

Prepared, autoclaved and stored as 2 separate solutions: phosphate buffer (Supplementary table 4) and 10x salts (Supplementary table 5).

Supplementary table 4: Composition of phosphate buffer for Burke's media

Quantity	**Component**
0.2 g	KH2PO4
0.8 g	K2HPO4
To a total of 900 mL	dH2O

Supplementary table 5: Composition of 10x salts for Burke's medium

Quantity	**Component**
200 g	sucrose
2.0 g	$MgSO_4 \cdot 7H_2O$
0.9 g	$CaCl_2 \cdot 2H_2O$

1mL	10mM $Na_2MoO_4 \cdot H_2O$ (aq)
50mg	$FeSO_4 \cdot 7H_2O$
	dH_2O up to 1 L

Every litre of B medium is then freshly prepared using 900 mL autoclaved phosphate buffer and 100 mL of autoclaved 10x salts solution. To prepare B medium plates, 16 g agar (Difco) is added to phosphate buffer prior to sterilisation. For BN medium, ammonium acetate was added to a final concentration of 15 mM.

9.2.2.2 BG11

Supplementary table 6: Composition of BG11 medium

Amount	**Component**
10 mL	100 x BG11 without Fe, phosphate, carbonate (see Supplementary table 8)
1 mL	1000 x Ferric ammonium citrate
1 mL	1000 x Na2CO3
1 mL	1000 x K2HPO4
To a total 1 L	dH_2O

Supplementary table 7: Composition of BG11 agar for solid media plates

Amount	**Component**
5 mL	100 x BG-11 without Fe, phosphate, carbonate (see Supplementary table 8)
0.5 mL	6 g/L Ferric ammonium citrate
0.5 mL	20 g/L Na_2CO_3
0.5 mL	30.5 g/L K_2HPO_4
5 mL	1 M TES/NaOH buffer pH 8.2
2.35 g	Pentahydrate Na-thiosulfate
7.5 g	Difco Bacto-agar
To a total 1 L	dH_2O

Supplementary table 8: Composition of BG11 without Fe, phosphate, carbonate

Amount	**Component**
149.6 g	$NaNO_3$
7.49 g	$MgSO_4 \cdot 7H_2O$
3.6 g	$CaCl_2 \cdot 2H_2O$
0.6 g 0.89 g	Citric Acid (or Na-citrate x $2H_2O$)
1.12 mL	0.25 M NaEDTA pH 8.0
100 mL	Trace minerals (see Supplementary table **9**)

Supplementary table 9: Composition of BG11 trace minerals	
Component	Amount
H_3BO_3	2.86 g
$MnCl_2 \cdot 4H_2O$	1.81 g
$ZnSO_4 \cdot 7H_2O$	0.22 g
$Na_2MoO_4 \cdot 2H_2O$	0.39 g
$CuSO_4 \cdot 5H_2O$	0.079 g
$Co(NO_3)_2 \cdot 6H_2O$	0.0494 g

9.2.3 Antibiotics Concentrations

Supplementary table 10: Antibiotics concentrations	
Antibiotics	**Final concentration (mg/ml)**
Spectinomycin	50
Carbenicillin	100
Ampicillin	100
Chloramphenicol	37
Kanamycin	50
Neomycin	10
Streptinomycin	20